Compilation of the

Construction Master Plan

DiSSCo Prepare WP9 – Deliverable 9.6

WP Leader

Eva Alonso (Naturalis Biodiversity Center) (ORCID: 0000-0001-5336-9723)

Authors

Naturalis Biodiversity Center, Leiden, The Netherlands

Tina Loo (ORCID: 0000-0002-8250-1786)

University of Florence, Natural History Museum, Italy

Marco Benvenuti (ORCID: 0000-0003-1344-8421)

Lorenzo Cecchi (ORCID: 0000-0001-9706-0947)

Gianna Innocenti (ORCID: 0000-0002-4504-0765)

Marta Biaggini (ORCID: 0000-0003-3968-9282)

Luca Bellucci (ORCID: 0000-0002-8464-6153)

Vanni Moggi Cecchi (ORCID: 0000-0001-5938-6276)

Fabio Di Vincenzo (ORCID: 0000-0002-2830-6971)

National Research Council of Italy - Institute of Geosciences and Earth Resources, Italy

Sarah Rossi de Gasperis (ORCID: 0000-0002-1609-7412)

Acknowledgements

The Authors would like to acknowledge the contributions of the many dedicated individuals in the DiSSCo Prepare family of experts without whose single-minded commitment to promise and principles, this Construction Master Plan would not have been possible. We gratefully thank the following individuals.

Stefaan Pijls   Meise Botanic Garden, BE

Julia Pim Reis   Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung, Berlin, Deutschland

Patrick Semal   Royal Belgian Institute of Natural Sciences, Brussels, BE

Serge Scory   Royal Belgian Institute of Natural Sciences, Brussels, BE

Vince Smith   Natural History Museum, UK

Sofoklis Sotiriou   R&D Department, Ellinogermaniki Agogi, Greece

Claus Weiland   Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt am Main, Germany

Matt Woodburn   Natural History Museum, UK

Contents

Abstract

The present document was developed under DiSSCo Prepare Project (DPP) Work Package 9 (WP9) "Project Management'', ensuring the synchronisation of project activities and final development of the DiSSCo Construction Master Plan (CMP). The CMP represents the compilation of all the outcomes of DPP and the collective effort and inputs of DiSSCo-related projects (SYNTHESYS+ and MOBILISE).

The Construction Master Plan (CMP) is a corpus of knowledge prepared as a design blueprint for the transition, construction and operation of the DiSSCo research infrastructure, structured to give immediate concrete and practical indications. It identifies over 40 targeted areas for construction supported by approximately 175 recommendations resulting from over 30 DPP deliverable outcomes. The CMP covers an extremely wide range of subjects commensurate with the scale and breadth of constructing a networked pan-European research infrastructure (RI) that adheres to the new principles and technologies inherent in OpenScience and FAIR data. For example, DPP deliverable outcomes report on legal entity formation; technology and e-services development (system architecture; data, information and knowledge systems); ensuring connectivity and capacity of consortium members; costing, cost recovery and contribution models; advocacy and engagement of members, partners, government ministries and other stakeholders; governance; policies and procedures; human resources and training, and how the RI can find its place in the international landscape of biodiversity research infrastructures.

The CMP was written first hand by project task leaders who provided concise summaries of task activities and outcomes, critically including prioritised lists of (actionable) recommendations and timelines to be implemented during the forthcoming 4-year period (2023-2026). The task summaries are the foundational basis for the CMP, while the T9.4 Team developed the Plan's framework and ensured quality, consistency and coherence essential for a sound blueprint to guide DiSSCo through its next transitional and construction phases.

The Distributed System of Scientific Collections (DiSSCo) is a pan-European research infrastructure committed to the digital unification of all European natural science assets under common access, curation, policies and practices, ensuring the principles of Open Science and FAIR (Findable, Accessible, Interoperable and Reusable) data.

01 Introduction (D9.6 DiSSCo Construction Master Plan)

1.1 Description

The present document represents the final outcome of the DiSSCo Prepare Project. It has been developed within Task 9.4 of DPP Work Package 9 (WP9) "Project Management", led by Naturalis Biodiversity Center (Leiden) with team members from the University of Firenze (Universitá degli Studi di Firenze), and the National Research Council of Italy - Institute of Geosciences and Earth Resources, Italy.

WP9 set up the governance and management structure of the DiSSCo Prepare project, ensuring the synchronisation of project activities. It was split into four tasks, focusing on the following key project-wide objectives:

· to support the operation and communication of the key governance and management structures of the project;

· to ensure day-to-day management of the project, by providing the right platforms for technical management, financial management and overall administration;

· to provide administrative and logistical support to the operation of the key advisory and stakeholders groups of the DiSSCo RI, as well as within the project team;

· to compile the DiSSCo Construction Master Plan (hereafter CMP) and to initiate (in collaboration with WP7 - Work Package 7 - Governance, Policy and Legal Framework) the process for establishing the new DiSSCo legal entity.

The CMP, i.e. the final outcome of Task 9.4 (deliverable 9.6) here reported, is a consistent corpus of knowledge, collating the outputs of all the work packages in DiSSCo Prepare and serving as guidance for the transition, construction, and eventual operation of the infrastructure.

1.2 Method

The Construction Master Plan, as well as the methodology used to develop it, is based on five core organisational elements (science, technical, data, organisation, and finances) set

forth in DiSSCo's Design Study (ICEDIG[1a]) and discussed in its summary output (Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1[2]). Whereas the remit of ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage) was to setup the structural pillars in the five core areas to enable operation of a unified bio- and geodiversity data access point, the DiSSCo Prepare Project (DPP[3]) raises implementation readiness levels across the five dimensions to ensure DiSSCo's ability to execute construction based on clear actionable guidelines with minimum risk. The final DPP deliverable compiles all outputs of its content related tasks to form the Construction Master Plan (CMP), the scientific, technical, organisational and financial blueprint for the formation of DiSSCo's legal entity, construction and eventual operation.

Developing the framework for a construction plan required a practical approach in identifying the different domains according to which the research infrastructure (RI) is expected to be implemented as well as the tangible elements to construct. Therefore a new outline for the construction plan was developed that captured all the tangible construction elements of DPP tasks and grouped them in a way that roughly approximates the original five core dimensions in concept but with slight modifications in naming and grouping. Specifically, the new outline was derived by breaking down DPP tasks into individual construction components based on their scopes and schedules (some DPP tasks were divided in two), ensuring representative names, and grouping related tasks into five chapters or functional systems (Data, Information and Knowledge; System Architecture; Governance, Strategy and Stakeholders; Finances; Human Resources). The outline of the CMP represents task information at three different levels of implementation: the DiSSCo project construction level, the chapter or system level, and the individual task level. This framework was reviewed and approved by Task and Work Package Leaders, and DiSSCo Coordination and Support Office (CSO) managers.

With the framework in hand, a method was devised to populate the Plan with DPP task outcomes written by Task Leaders themselves. Task summary forms were developed and shared on a server with Task Leaders requesting them to provide a task description, methods, discussion of results, prioritised actionable recommendations, an approximate timeline for implementation, and identification of any gaps. To ensure that tasks still needing further attention or work during pre- or post-construction periods were captured in the CMP, the period covered by the CMP was expanded on both ends to include the Transition and first possible year of Operation (2023-2026). The content of the Task Summaries were then transferred to the CMP to form its textual basis. If tasks were not yet completed, the forms were populated with provisional information or outcomes, and updated at completion.

Multiple members of the T9.4 team reviewed each of the 41 completed task summaries providing editorial comments to the author ensuring completeness, comprehensibility,

coherence, and consistency with other chapter tasks. Although team members were not expected to be experts in the diverse array of DPP subjects, their review was essential to the integrity and quality of the document. Meetings were arranged with some task leaders when required, otherwise most issues were resolved via requested revisions or exchanges with the author via comments and responses in the shared document.

Near the end of the DPP project, completed task summaries were transferred to the CMP deliverable template. Task Summary text was directly copy/pasted into the template, while all DPP Task recommendations were compiled in a separate spreadsheet. The spreadsheet was used as a Master file to filter, sort by priority, and then transfer custom data sets to the deliverable template depending on the CMP subject. The recommendations were aggregated at the three different levels of construction implementation (project, chapter or system, and task) and represented as tables (see below).

a) DiSSCo Construct (All Tasks)

NB: ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

Sector	Task Name	Pri Typ	Rec No,	Recommendation	2023				2024				2025				2026			
				"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
Fin	Cost Book	Tr	137	**(H) Design a training for cost estimation.** Should be designed before construction phase and implemented in 2024 ⇒ NN representatives fully trained by 2026	P	P	I	I	I	I	I	F	F							
HR	Training Strategy	Tr	155	The Training Strategy for DiSSCo constitutes the starting point to first develop and then implement the DiSSCo Training Programme. It provides substantiated information about the existing capacities and the detected demand across the community that needs to be further elaborated. During the Transition phase, specific actions might be required: **Embed the DiSSCo Training into the CETAF-DEST and articulate how to implement it, following the recommendations from DPP.**		P	P	I	F											
HR	Training Strategy	Tr	156	During the Transition phase, specific actions might be required: **Deploy the DiSSCo training service into annual training programmes in accordance to the priorities detected to ensure DiSSCo operation across institutions as providers of data.**				P	I	F										
HR	Training Strategy	Tr	157	During the Transition phase, specific actions might be required: **Align the DiSSCo programmes with other complementary resources to avoid duplication of efforts (such as GBIF) by subscribing specific agreements for collaboration.**		P	P	P	I	F										
DIK	Digitisation Prioritisation (Institutional and DiSSCo)	H	001	**We recommend collaboration on digitisation projects,** especially at DiSSCo level, but the criteria for prioritisation are also applicable to institution-specific, national or regional projects. Coordination among DiSSCo partners will support the community itself and the rapid developments in approaches which are happening around the world as a solution to help drive forward strategic prioritisation of digitisation activities. Communicating summaries of these and adding to these will have a dual role in helping others define or refine their strategies.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

2) Governance, Strategy and Stakeholders

NB: ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Task Name	Pri Typ	Rec No.	Recommendation	"P" for "Planning", "I" for "Implementation", "F" for "Finalisation", "O" for Ongoing															
				2023				2024				2025				2026			
				1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
Governance & Operational Planning	Tr	104	**ERIC Statutes and financial provisions approval**	I	F														
Governance & Operational Planning	Tr	105	**Launch Call for Hosting Country and Founding Members**	P	I	I	F												
Governance & Operational Planning	Tr	106	**Prepare Technical and Scientific Descriptions** of the ERIC	P/I	F														
Governance & Operational Planning	Tr	107	**Prepare Step 1 and submit**			I	F												
Governance & Operational Planning	Tr	108	**Prepare bylaws and operational rules**	P	I	I	F												
Governance Strategy & Strategic Implementation	Tr	113	**Identify strategic actions for construction phase and first years of operation. (H)**	I	F														
Legal Entity (ERIC statutes, sci/tech desc, Step1)	Tr	116	**Keep the DiSSCo Community active** (through regular National Nodes meetings and iGAs) **to support their efforts for motivating their national authorities** to have DiSSCo ERIC established.	I	I	I	I	F	F										
Legal Entity (ERIC statutes, sci/tech desc, Step1)	Tr	117	**Get formal approval of the Statutes and bylaws by the (representatives of the) countries willing to support DiSSCo ERIC** from the beginning.	I	I	I	I	F	F										

c) Pre-commercial Procurement

NB: ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri Typ	Rec No,	Recommendation	"P" for "Planning", "I" for "Implementation", "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
			1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
H	145	**Lock the DiSSCo Construct development needs (H, prep needed)**: By listing the development needs according to their industry TRL' s and their fit with the 4 strategic pillars defined during DiSSCo Prepare, we can create a list of eligible activities and their priorities, that can then serve as pipeline for PCP and PPI opportunities. This should be finalised during the remaining time in DiSSCo Prepare.	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F
H	146	**Establish DiSSCo as a legal entity (H, prep needed)**: In order to leverage the benefits of PCP and PPI during the Construct Phase of DiSSCo, we need a legal entity or one of the consortium partners to assume the procurement role during the construct phase. At which time, we should also conclude on the procurement structure in DiSSCo	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F
H	147	**Lock on DiSSCo procurement structure (H, prep needed)**: the procurement function can be centralised, decentralised, hybrid,.. Each setup has its own merits and drawbacks, nevertheless, a timely choice would benefit the future efforts in DiSSCo.	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F

Figures 1a, 1b, 1c. Prioritised task recommendations and timeline from DPP compiled at three levels of project implementation: a) project level, b) chapter (or 'system') level, c) task level.

The Draft CMP with integrated text and tables was shared with Task/WP Leaders and CSO Managers. Task and WP Leaders were asked to update their text and recommendations

with recently finalised outcomes, and CSO Managers were asked to provide an introduction and overview of each chapter integrating major task summary outcomes and dependencies, address implementation readiness levels, and make relevant projections for construction including gaps.

The finalised document was reviewed by CSO Management before submission.

1.3 Structure of the CMP

The CMP is a guideline for the next steps in DiSSCo's transition and construction, leading to the operation of the RI. It is structured to provide immediate concrete and practical indications. In this regard, a first chapter provides a general overview of the entire DiSSCo Construct project followed by a complete list of DPP recommendations and their timelines in table format. The recommendations here are listed by priority, regardless of the chapter and task they belong to.

The following chapters as stated, are grouped by related tasks into five chapters or functional systems (Data, Information and Knowledge; System Architecture; Governance, Strategy and Stakeholders; Finances; Human Resources). Each chapter is preceded by a synoptical graphic representing interdependencies existing between tasks in the chapter with other tasks or chapters (systems). The structure of each chapter is similar and includes a general introduction written by experts of the DiSSCo CSO, followed by chapter (system) prioritised recommendations, its related individual task summaries and their prioritised recommendations. The multi-level representation of the tasks, from the individual, to system, to project level facilitates the required construction perspectives.

Please note: For full functionality with document links, it is recommended to use the Word document in Word, or the pdf document in Adobe. You can navigate quickly through

- the **Word document in Word** using the **Contents** (CTRL Click on the desired section and ALT left arrow to return); using **Word's Navigation Pane** (CTRL 'F' or View/Navigation Pane); or, **CTRL click on the links** throughout the document.

- the **PDF document in Adobe**, Click or CTRL Click on the link to navigate to it; return with ALT left arrow.

1.4 Tasks 1.1 and 1.2, Life Sciences and Earth Sciences Use Cases and User Stories

These two DPP tasks have not been included in the CMP in the standard manner as other tasks because they do not have a construction component. However, their output, the

collecting and analysing of DiSSCo user stories, is foundational to the design and construction of the research infrastructure and is thus summarised below.

The planned DiSSCo Research Infrastructure (RI) will be an important source of information for scientists from natural science disciplines but also other users from the sectors education, culture, society, politics, and economy. In order to meet the requirements of all potential stakeholders, the planning and construction of the DiSSCo RI is strongly user driven. DiSSCo Prepare Tasks 1.1[4] and 1.2[5] examined the needs of different stakeholder groups for the information that natural science specimens and collections contain and the requirements these needs set for the services to be provided by DiSSCo. More closely, the Tasks 1.1 "Analyse Life sciences use cases and user stories" and 1.2 "Analyse Earth sciences use cases and user stories", built on existing studies and compilations covering DiSSCo-related use cases and user stories. Task 1.1 and Task 1.2 were complementary to each other, focussing on the two domains Life sciences and Earth sciences, respectively. While Task 1.1 dealt with biological collections (entomological, other zoological, botanical and mycological collections), the focus of Task 1.2 was on collections of fossils, rocks, sediment structures, minerals, and extra-terrestrial material (meteorites). The complemented corpus of Life sciences and Earth sciences user stories and use cases was analysed with a special emphasis on the functional demands for DiSSCo and its services, as well as their socio-economic importance.

The use cases were grouped into the seven user groups or use categories, which were also adopted from the ICEDIG project (van Egmond et al. 2019):

1. Research (academic, non-academic incl. Citizen Science)
2. Collection management
3. Technical support (IT & IM)
4. Policy (institutional, national & international)
5. Education (academic & non-academic)
6. Industry
7. External (media & empowerment initiatives)

By far, the most use cases were represented in "Research" and "Collection management". The societally wide-ranging needs for the use of scientific collections came to the fore. For instance high-quality metadata descriptions and images are highly needed and serve stakeholders from research to industry. The recognition and description of functional demands at appropriate and useful levels required several rounds of refinement to optimise usefulness for the further development within DiSSCo. The analysis of use cases and recognition of functional demands creates a basis for and supports further RI DiSSCo development, e.g. recognition of digitisation prioritisation criteria and set the service development framework.

All identified use cases[6] were imported into GitHub, the main repository for technical developments in DiSSCo Prepare and related projects (DiSSCo/user-stories). Therewith, they are easily accessible for development teams and can be taken into consideration when setting up development plans for DiSSCo technical architecture, the prioritisation of services and setting up or shaping pilots in DiSSCo Prepare.

02 DiSSCo Transition & Construct

2.1 Overview

The development of a European Research Infrastructure in the context of the ESFRI (European Strategy Forum on Research Infrastructures), typically goes through a series of distinct, yet highly interdependent, phases. The most prominent ones are the Preparatory Phase, the Transition Phase, the Construction Phase and the Operational Phase. With the closure of the DiSSCo-linked projects (SYNTHESYS+, DiSSCo Prepare and MOBILISE), DiSSCo is successfully wrapping up its Preparatory Phase (2018-2023). In this document, we summarise the key outcomes of this Phase and provide actionable recommendations for the next Phase of the infrastructure development cycle. During the Preparatory Phase we organised our work around five distinct dimensions, aiming at improving the overall Implementation Readiness Level (IRL) of the infrastructure. Significant progress was made in improving our (i) data, (ii) technology, (iii) financial, (iv) organisational and (v) legal capacity. We advanced our joint understanding of the fundamentals of constructing and operating a complex distributed infrastructure, whilst we identified the areas where more work is needed. Finally, we improved the links between science and technology, by extensively and systematically describing use cases and using them to understand the functional requirements of the future DiSSCo tools and services.

The Transition Phase of DiSSCo will start in 2023 and immediately after the closure of the Preparatory Phase projects. The main goal of this Phase is to set-up the legal entity of DiSSCo (in the form of an ERIC-European Research Infrastructure Consortium). At the same time, however, during the Transition Phase it is essential that the infrastructure a) addresses issues identified during the Preparatory Phase, b) continues the development of its core technical infrastructure and c) continues to enhance the associated social infrastructure (e.g. community engagement and user skills development).

The Construction Phase of DiSSCo will commence at the time when its legal entity will be in place together with its new funding model (member states membership fees). We anticipate this new Phase to start in Q1 of 2025. At that time DiSSCo ERIC will implement a short and highly intensive project with the overall aim to deliver its full service portfolio within two years (2025-2026). This project (DiSSCo Construct) will focus mainly on operationalising the portfolio of e-services designed and piloted during the Preparatory

Phase, connect technical components in a way that provides an integrated view into the data and applications mobilised and maintained by DiSSCo, and finally formalise the agreements needed between DiSSCo ERIC and its distributed service providers (National Nodes and other organisations).

Despite the distinct nature (different funding sources and work intensity) of the two Phases above, we address the needs of the infrastructure between 2023 until 2026 as a continuous development programme. The priorities of this programme will be informed by the results of the community efforts, which are presented in a structured and systematic way and in the form of thematic recommendations, in this document.

2.2 Timeline and Summary Recommendations

Table 1. DiSSCo Transition and Construct (All Tasks) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

Sector	Task Name	Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
					2023				2024				2025				2026			
					1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
DIK	3.3.2 Data Management Plan Implementation (ICEDIG))	GAP	010	"Machine-actionability", the capability of machines to handle data autonomously and appropriately is a core objective of FAIR and thus FDOs (Jacobsen 2020[7a]). The concept was i.a. developed to describe an aspired transformation of DMPs from essentially static documents towards an application-based tool chain integrating and automating all relevant information steps in the data life cycle: the machine-actionable data management plan (maDMP; Miksa 2019[8a]). **Due to several constraints in DPP, a merge of the maDMP concepts and the FDO approach wasn't elaborated to the same extent as a prototype in DPP like the machine-learning and human- (crowd-) based annotation pipelines** (cp. Leeflang 2022[9a]; Grieb 2021[10a]). The objective and potential of the maDMP are the seamless integration of the reporting tools into the																

existing network of project tools and services to release involved humans (developers, scientists, curators etc) from the "annoying administrative exercise" (Miksa 2019) to write a DMP as a matter of form. It should be feasible to realise this level of automatisation for DiSSCo's Digital Specimen services, since the fundamental Digital Object Architecture provides already a network of FAIR-enabling services, which (prototypically) implement the policies, rules, procedures and infrastructures for specimen data during its whole life cycle. **The main task now is to provide a proper alignment on several levels (data model, services) with common specifications (DCAT, PROV-O) to make the DES more "speech-enabled" with regard to retrospective and prospective provenance.**

DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	Tr	004	**Further develop and refine the provenance model in openDS.**	I	I	I	I													
DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	Tr	005	**Provide a more complete capture of digitisation processes in MIDS** by establishing a DES as output of a MIDS level 3 description (align MIDS information elements with openDS) with the aim to facilitate the transfer of digitisation workflows.	I	I	I	I													
SysArch	4.3.1 Implementation of the Digital Specimen	Tr	043	(Before 2024): **Finalisation of a first release of the openDS specification** (see section "Discussion and Conclusions" for details). Immediate goal is a first release within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

	Architecture (DS Arch)																			
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	Tr	044	(Before 2024): **Develop a coherent digitisation strategy binding for DiSSCo partners** including set up of infrastructures like sufficient storage capacities and the implementation of specifications (openDS, MIDS) to enable reliable feed into DiSSCo's core data pipelines.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.4 Integration in the Global Technical Landscape	Tr	093	**Finalise the first release of the openDS specification** (a technical necessity, see section "Discussion and Conclusions" in CMP T6.4[11a] for details) within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.4 Integration in the Global Technical Landscape	Tr	094	**Ensure close involvement and alignment of DS/DS architecture with developments of Digital Object Architecture, FAIR Digital Objects Forum (FDOF) FDOs including FDO typing, machine actionability, and FDO profiles and attributes.** As noted above, the development of DiSSCo's Digital Specimen concept involving the corresponding service architecture is highly leveraging on Digital Object Architecture. DOA and in particular FDOs as core data models are currently the subject of intensive review and (re-)specification process. The procedure is documented in the FDOF Public Document Management Register[12a]). Of particular importance for the DS/DS architecture are the specifications of FDO typing (important with regard to subclassing DS, see subclasses proposed in Addink 2022 like BotanySpecimen, MycologySpecimen, GeneticResourcePlantSpecimen etc.),	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

				Machine-Actionability (the core objective of the FAIR principles to realise a global data ecosystem that can be navigated independently by machines), and FDO PID Profiles and Attributes (which provides essential Kernel Attributes of the FDO when an FDO's PID is resolved, cp. Islam 2020[13a]). It is fundamental for DiSSCo's technical architecture to be fully involved in this process and to ensure a close alignment with delivered results. Currently, members of DiSSCo's technical team are involved in the drafting process of all aforementioned documents with particular relevance for DiSSCo DS architecture.																					
GSS	5.3.1 Governance and Operational Planning	Tr	104	**Statutes and financial provisions approval (ERIC)**	I	F																		
GSS	5.3.1 Governance and Operational Planning	Tr	105	**Launch Call for Hosting Country and Founding Members (ERIC)**	P	I	I	F																
GSS	5.3.1 Governance and Operational Planning	Tr	106	**Prepare Technical and Scientific Descriptions (ERIC)**	P/I	F																		
GSS	5.3.1 Governance and Operational Planning	Tr	107	**Prepare Step 1 and submit (ERIC)**			I	F																

GSS	5.3.1 Governance and Operational Planning	Tr	108	**Prepare bylaws and operational rules (ERIC)**	P	I	I	F																					
GSS	5.3.2 Governance Strategy and Strategic Implementation	Tr	113	**Identify strategic actions for construction phase and first years of operation. (H)**	I	F																							
GSS	5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1)	Tr	116	**Keep the DiSSCo Community active** (through regular National Nodes meetings and iGAs) **to support their efforts for motivating their national authorities** to have DiSSCo ERIC established.	I	I	I	I	F	F																			
GSS	5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1)	Tr	117	**Get formal approval of the Statutes and bylaws by the (representatives of the) countries willing to support DiSSCo ERIC** from the beginning.	I	I	I	I	F	F																			
GSS	5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical	Tr	118	**Enter and complete the two-steps submission procedure to the EU Commission.**	I	I	I	I	F	F																			

				Description																					
	Description, Step 1)																								
GSS	5.3.5 National Nodes Engagement	Tr	121	**(H) Continue supporting the node's involvement in the building process.** During the transition phase, CETAF will act as a mechanism to secure involvement of the NNs and obtain their feedback and contributions; During the Construction phase, once the ERIC is in place, the participation of the nodes will be secured with the Nodes Committee, an Advisory Body to the GA. Please note: In the timeline only the Transition phase is covered. For the Construction phase, the Nodes committee shall be implemented under the ERIC governance model.	P	I	I	I	F																
GSS	5.3.5 National Nodes Engagement	Tr	122	**(M) Continue to launch surveys and compile information at node level for critical aspects of the DiSSCo architecture** and also for other capacity services (as training), as it remains critical for delivering well-grounded reports and driving good evidence-supported conclusions. To be meaningful, the information collected cannot be exclusively from large institutions but should cover the entire community.	P	I	I	I	F																
GSS	5.3.5 National Nodes Engagement	Tr	123	**(M) Foster and enlarge participation across countries with periodic updates from the nodes** on the national priorities and the integration of digitisation activities in the national roadmaps.	P	P	P	I	I	I	I	I	I	I	I	F							
GSS	5.3.5 National Nodes Engagement	Tr	124	**(H) Encourage the adoption at national level of relevant DiSSCo services and developments** such as ELViS and support the population of relevant components with data (such as the Collections Registry as the feeding pipeline	P	I	I	I	F																

Group	Task	Type	No.	Description	1	2	3	4	5	6	7	8	9
				to ELViS). **The compilation of Collections Registry data should be finalised before construction and the linkage with aggregators (such as GRScicoll) shall be formalised during the transition phase to avoid duplication and lack of interoperability.**									
GSS	5.3.5 National Nodes Engagement	Tr	125	**(M) Complete the Specialization Plan before 2024** to serve as a vehicle for the CSO (and once constituted, to the ERIC) to better understand participation at national level.	P	I	I	I	F				
Fin	6.3.1 Cost Book	Tr	137	**(H) Design a training for cost estimation.** Should be designed before construction phase and implemented in 2024 ⇒ NN representatives fully trained by 2026.	P	P	I	I	I	I	I	F	F
HR	7.3.2 Training Strategy	Tr	165	The Training Strategy for DiSSCo constitutes the starting point to first develop and then implement the DiSSCo Training Programme. It provides substantiated information about the existing capacities and the detected demand across the community that needs to be further elaborated. During the Transition phase, specific actions might be required: **Embed the DiSSCo Training into the CETAF-DEST and articulate how to implement it, following the recommendations from DPP.**		P	P	I	F				
HR	7.3.2 Training Strategy	Tr	166	During the Transition phase, specific actions might be required: **Deploy the DiSSCo training service into annual training programmes in accordance to the priorities detected to ensure DiSSCo operation across institutions as providers of data.**				P	I	F			
HR	7.3.2 Training Strategy	Tr	167	During the Transition phase, specific actions might be required: **Align the DiSSCo programmes with other**		P	P	P	I	F			

				complementary resources to avoid duplication of efforts (such as GBIF) by subscribing specific agreements for collaboration.																	
DIK	3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)	H	001	**We recommend collaboration on digitisation projects,** especially at DiSSCo level, but the criteria for prioritisation are also applicable to institution-specific, national or regional projects. Coordination among DiSSCo partners will support the community itself and the rapid developments in approaches which are happening around the world as a solution to help drive forward strategic prioritisation of digitisation activities. Communicating summaries of these and adding to these will have a dual role in helping others define or refine their strategies.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
DIK	3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)	H	002	**We suggest that DiSSCo and its partners apply the guidelines for prioritisation of digitisation developed by Task 1.3,** which suggest to **focus on the four main criteria established, i.e. Relevance, Data quality, Cost and Feasibility, and to answer the relevant questions that were provided.**	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
DIK	3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)	H	003	**For the evaluation process we recommend a combination of a scoring system and a panel review,** similar to how applications have been evaluated in the SYNTHESYS projects[14a]. We consider a panel review to be necessary because of the diversity and multi-dimensionality of the prioritisation criteria. We are aware that the guidelines may need modification according to future experience gained.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

				Description																		
DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	H	006	**Establish a FDO type for a machine-actionable DMP** Link all relevant outputs of a DES' lifecycle to a maDMP-FDO comprising links to retrospective and where applicable prospective provenance objects, to related services like re3data for repository information, to machine-actionable policies in a structured format like ODRL. Elements of this FDO type must support a sufficient level of granularity (no long free text as bitstream) to enable reaggregation of information elements to adjust the DMP dynamically.	P	P	P	P	I	I	I	I	F	F	F	F						
DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	H	007	**Implement linking to comparable service networks like ENVRI-FAIR** Enable the interoperability of DiSSCo's core entities involved in RDM by providing mappings to a common vocabulary shared among relevant data providers like the ENVRI-FAIR network. A common standard here is W3C DCAT, which is centered on requirements of data exchange of data portals in Europe providing semantic interoperability by reuse of (e.g. EuroVoc) and mappings to existing vocabularies (e.g. INSPIRE, schema.org[15a]). Predefined profiles are available like DCAT-AP used for EU infrastructures, EPOS DCAT-AP employed in ENVRI-FAIR which provides starting points for DiSSCO's mapping (or subclassing) of core entities like DES, collection, CMS, enrichment service to dcat:Resource, dcat:catalog dcat:Service etc.	P	P	P	P	I	I	I	I	F	F	F	F						
DIK	3.3.3 Digital Specimen Data Model	H	011	We foresee that the basic modelling processes will have been established with the end of the DiSSCo preparation phase. However, there will be the need for further	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F	F		

				coordination of works on the model itself. The DiSSCo synchronisation group for data standards strongly recommended allocating resources for the coordination of modelling activities. **Allocate resources for the coordination of further modelling activities and recruit a modelling expert.**																
DIK	3.3.4 Semantic Annotation	H	013	**Develop tools that enable automated enrichment** both as part of the digitisation process and as part of publication pipelines.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
DIK	3.3.4 Semantic Annotation	H	014	**Organise a working group to exchange experiences on enrichment, and discuss and promote curation of open resources** such as Wikidata to avoid double-work.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
DIK	3.3.5 Data Mobilisation (Institutional)	H	019	**Establish a Digitisation Interest Group** with representation and linkage to other relevant community groups (e.g., TDWG, CETAF DWG/ISTC) to continue development of the DiSSCo Digitisation Guides, and discuss challenges and solutions for all aspects of digitisation.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
DIK	3.3.5 Data Mobilisation (Institutional)	H	022	**Determine how much, if any, investment DiSSCo plans to make in digitisation infrastructure and how much support will be provided to institutions to increase their digitisation capacity.** This includes support to keep the Digitisation Guides up-to-date and to foster community curation. This would ideally be combined with both in-person and online training, and potentially support planning and implementation of physical and digital infrastructure investment. We recommend focusing content and support on new or recently established	P	P	P	P	I	I	F	F	O	O	O	O	O	O	O	O

				national nodes where digitisation is a priority, and national nodes which are working on national or regional digitisation programmes.																		
DIK	3.3.6 Taxonomic Backbone	H	023	**Focus on articulating more detailed requirements for taxonomic names services** *Species 2000 would recommend a further in-depth exploration with DiSSCo into the requirements for taxonomic names services. *Engage with COL to further explore the benefits that COL and ChecklistBank might leverage from interoperability with DiSSCo, and resource implications of supporting those requirements in the DiSSCo architecture.	P	I	F														
DIK	3.3.6 Taxonomic Backbone	H	024	**Assess and refine the potential technical approaches to interoperability with ChecklistBank** *Use technical pilots to assess the possibilities and limitations of using COL's public services and open APIs to meet DiSSCo integration requirements. *Engage with COL to explore the potential implementation of the event-based interoperability model, and any associated investment that might be required into COL development.	P	I	I	I	F											
DIK	3.3.7 Knowledgebase	H	027	**Establish a dedicated person (editor) or editorial team** responsible for quality control, content management and curation of the DiSSCo KB.		P	P	P	I	I	I	I	I	F	F	O	O	O	O
DIK	3.3.7 Knowledgebase	H	028	**Engage or provide technical maintenance services and basic software support** for KB.		P	P	P	I	I	I	I	I	F	F	O	O	O	O
DIK	3.3.7 Knowledgebase	H	029	**Further, integrate with DiSSCo's <u>Authentication and Authorization Infrastructure</u>** (AAI).		P	P	P	I	I	I	I	I	F	F	O	O	O	O

DIK	3.3.7 Knowledgebase	H	030	**Further, implement permanent ongoing curation, maintenance and development** (based on future requirements) of the service to remain a trusted information source.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
DIK	3.3.8 Helpdesk	H	035	**Offer three complementary services for DiSSCo Helpdesk: FAQs - Frequently Asked Questions; Human Intermediated Question and Answer Service; and Supporting Documentation.** With respect to supporting documentation, most DiSSCo services are still under development and, therefore, production of supporting documentation should be considered an ongoing process.	P	P	P	P	I	I	I	I	I	F	F				
DIK	3.3.8 Helpdesk	H	036	**For DiSSCo Helpdesk use the JitBit Helpdesk Ticketing System adopted for ELViS helpdesk.** Considering that the DiSSCo community is already familiar with this system, we anticipate no major issues, and no further preparation activities are needed to initiate its construction in 2024. We are confident that its construction will be finished by 2025.	P	P	P	P	I	I	I	I	I	F	F				
DIK	3.3.8 Helpdesk	H	037	**Integrate DiSSCo Knowledgebase (KB) with DiSSCo Helpdesk** which will provide links to new and/or relevant content; automatically generate search results based on incoming questions, supporting personalised responses; and linking to other DiSSCo eServices, etc.	P	P	P	P	I	I	I	I	I	F	F				
DIK	3.3.8 Helpdesk	H	038	**Follow the recommendations of the Web Accessibility Initiative (WAI) for the Helpdesk,** embracing an inclusive design that meets the needs of those who are permanently or temporarily disabled.	P	P	P	P	I	I	I	I	I	F	F				

	Section		#	Description																
DIK	3.3.8 Helpdesk	H	039	Allow real-time monitoring of the support process in The Helpdesk, namely the Human Intermediated Q&A Service. **Develop a Service Level Agreement** (SLA), in which goals are set to meet users' expectations.	P	P	P	P	I	I	I	I	F	F						
DIK	3.3.8 Helpdesk	H	040	**Provide** training **to support staff on policies and procedures, as well as on the tools and technology** they will be using. This will help ensure that the Helpdesk team is able to provide high-quality support to users.	P	P	P	P	I	I	I	I	F	F						
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	045	**Transfer pilot plans of DiSSCo's service hub (e.g.** repo[16a]**, data pipelines) detailed in D6.2 Implementation and construction plan to full production facilities** (Leeflang 2022[17a]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	046	**Develop, coordinate and operate a distributed, cross-institutional service support network to facilitate the local implementation of necessary technical adjustments linked to the Digital Specimen architecture.**	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	047	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation **[H] (concept: "UCAS - Unified Curation and Annotation System")** and annotation [M-L] (prototypes Specimen Data Refinery Hardisty 2022[18a], Machine Learning as a Service (MLaaS) pilot Grieb 2021[10b]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

SysArch	4.3.2 Interoperability (Institutional CMSs)	H	051	**Install a local dedicated CMS person (developer) responsible for development to implement specific response procedures triggered by new event types.** For small to medium sized museums, we recommend hiring a developer at least for the first three months or for the period agreed upon between the institution and the professional to do the implementation requested by this event-driven API on the local CMS.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.2 Interoperability (Institutional CMSs)	H	052	**Install a local CMS technical team for ensuring compatibility** while both the CMSs and DiSSCo RI evolve.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.2 Interoperability (Institutional CMSs)	H	053	**Provide technical maintenance services and basic software support.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.2 Interoperability (Institutional CMSs)	H	054	**Further, develop, coordinate and operate a cross-institutional DiSSCo CMS support network (DiCSN).** A DiCSN is required to embed local CMS infrastructures of participating institutions in DiSSCo's Digital Specimen architecture. This will be based on the implementation of the API guidelines detailed in DPP D6.1 Harmonisation and migration plan for the integration of CMSs into the coherent DiSSCo Research Infrastructure (Glöecker et al. 2022[19a]) in close alignment with further development of the ingestion data pipelines outlined in DPP D6.2 Implementation and construction plan of the DiSSCo core architecture (Leeflang 2022[17b]). As a major	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

					T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
				outcome of D6.1, detailed API guidelines for the bi-directional interfaces between local CMSs and the DiSSCo core infrastructure were elaborated. These guidelines could be either implemented as an extension of existing application interfaces to set up DiSSCo specific endpoints in CMSs that are web-based or by provision of a light-weight wrapper, a technique for extracting the data from documents that belong to non-web-based CMSs. To keep the technical integration of the local CMSs synchronised with the ongoing development of the retrieval infrastructure is of major importance to (i) enable continuous data feed into the core architecture (translator services, see D6.2) and to (ii) ensure the reverse data flow back from distributed curation into the local CMSs.																		
SysArch	4.3.3 Technical Services Development and Interoperability with DS Arch	H	058	**Find maintainers for the core e-services and plan steps to transform into future DiSSCo service providers.**	P	P	P	I	I	I	I	I	I	F	F	F						
SysArch	4.3.3 Technical Services Development and Interoperability with DS Arch	H	059	**Develop the e-services into mature products for operation and integrate them with the DiSSCo FAIRDO infrastructure**	P	P	P	I	I	I	I	I	I	F	F	F						
SysArch	4.3.3.a Digital Specimen Repository	H	062	**Implement Digital Specimen Repo as a component within the DiSSCo Arch.** Tracked in DD-132[20a].	P	P	P,I	P,I	P,I	F												

	(DSR) Infrastructure																						
SysArch	4.3.3.a Digital Specimen Repository (DSR) Infrastructure	H	063	**Adopt the openDS data model.** Tracked in DD-133[21a].		P	P	P,I	P,I	F	F												
SysArch	4.3.3.b Data Processing and Publishing Infrastructure	H	066	**Ensure data processing and publishing functionality** as part of the deployment of the Digital Specimen Repository (tracked in DD-132[20b])	P	P,I	P,I	P,I	F														
SysArch	4.3.3.c Persistent Identifier (PID) Infrastructure	H	067	**Build a roadmap for DOI implementation** (from Handle to DOI) Tracked in DD-5[22a].		P	P	P	P,I	P,I	F												
SysArch	4.3.3.c Persistent Identifier (PID) Infrastructure	H	068	**Create a service management plan for the PID service.** Tracked in DD-113[23a].	P	P	P,I	P,I	F														
SysArch	4.3.3.d Authentication and Authorization Infrastructure (AAI)	H	070	**Finish User Profile development work and deploy it** in the current sandbox. Tracked in DD-13[24a].	P,I	F																	
SysArch	4.3.3.e Indexing and API Service	H	074	**Ensure API functionality** as part of the deployment of the Digital Specimen Repository[25a] (tracked in **DD-132**[20c]).	P	P	P,I	P,I	F														

	Components Infrastructure																							
SysArch	4.3.3.e Indexing and API Service Components Infrastructure	H	075	**Provide API documentation.** Tracked in DD-135[26a].			P	P,I	P,I	F														
SysArch	4.3.3.f Unified Curation and Annotation System	H	076	**Finish UCAS Enhancement**. Tracked in DD-26[27a].	P,I	P,I	F																	
SysArch	4.3.3.f Unified Curation and Annotation System	H	077	**Consult users and redesign with UI/UX expert.** Tracked in DD-18[28a].		P	P	I	F															
SysArch	4.3.3.h European Loans and Visits System (ELViS)	H	083	**Fix and improve current ELViS code base** (ELViS Code Evaluation[29a]). The Epic (DD-16[30a]) (as of Dec 2022) has several tickets in different states of completion.	I	F																		
SysArch	4.3.3.h European Loans and Visits System (ELViS)	H	084	**Align and integrate ELViS with the DiSSCo core infrastructure development** activities **and data management plan.** Also tracked in (DD-16[30b]).	P	I	I	F																
SysArch	4.3.3.i Specimen Data Refinery (SDR)	H	087	**Align SDR with the DiSSCo core infrastructure development** activities. Tracked in DD-6[31a].	P,I	P,I	P,F																	

Domain	Task	Pr	No	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SysArch	4.3.3.j Collections Digitisation Dashboard (CDD)	H	090	**Automate import of digitised and collection holdings data into CDD** (note: this involves more effort from the collection holdings institutes than the DiSSCo technical team, and will therefore depend on the local technical team their resources and training).	P	Pj	P,I	P,I	I,F	I,F	I,F	I,F	I,F	I,F	I,F					
SysArch	4.3.3.j Collections Digitisation Dashboard (CDD)	H	091	**Align CDD with TDWG standard** (Latimer Core[32a], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). This is related to DD-119[33a] and DD-32[34a].			P	P	P	I	I	I	F	F	F					
SysArch	4.3.4 Integration in the Global Technical Landscape	H	095	**Develop the infrastructures for a FAIR service ecosystem leveraging FAIR Digital Objects for the greater range of Biodiversity, Life and Earth System Sciences. Provide the ecosystem of key services for FAIR Digital Objects** (Collins 2018[35a]) like FDO type and operation (for machine agents) registries, services for minting PIDs (cp. Hardisty 2021[36a]), repositories for the FDOs not only for the narrower target group of DiSSCo (NSCs) but also for the greater range of Biodiversity, Life and Earth System Sciences. The fundamental service ecosystem will be supplemented by community services like tools for modelling of new FDO types (for the diversification of DS subtypes see Addink 2022[37a]) or terminology services (TS) for the curation of semantic artefacts, including aforementioned mappings e.g. of DS concepts to classes of the new GBIF data model (a blueprint for a TS embedded in a FDO layer was drafted	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

Team	Ref	Pri	No	Description													
				with participation of DiSSCo as Flexible Semantic Mapping Framework/SEMAF; Broeder 2021[38a]).													
SysArch	4.3.5 Geo-collections Mobilisation	H	098	In collaboration with TalTech, MfN, the GeoCASe Advisory Board and the CETAF Earth Science Group, **continue to develop the joint vision for the development of GeoCASe,** explore models and opportunities for funding and resourcing GeoCASe development, and assess the overlap with GBIF in the palaeontology domain.	P	I	I	F									
SysArch	4.3.5 Geo-collections Mobilisation	H	099	**Focus initially on publishing data to GeoCASe** through the preparation and presentation of Darwin Core Archives that GeoCASe can ingest.				P	I	F							
SysArch	4.3.5 Geo-collections Mobilisation	H	100	**Incorporate support for Mindat classification and locality identifiers into the openDS specification.**				P	I	F							
GSS	5.3.1 Governance and Operational Planning	H	109	**Prepare Step 2 (ERIC)**		P	P	P	P	P	F						
GSS	5.3.1 Governance and Operational Planning	H	110	**Constitute ERIC**							F						
GSS	5.3.1 Governance and Operational Planning	H	111	**Keep principles of transparency, flexibility and agile decision-making process as pillars for governance** during operation.										O	O	O	O

Team	Section	Pri	#	Task																		
GSS	5.3.1 Governance and Operational Planning	H	112	**Use an inclusive methodology (Design Thinking) standard for all preparatory works** concerning ERIC application and operational planning.															O	O	O	O
Fin	6.3.1 Cost Book	H	138	**Develop a web platform to ease cost assessment.** Can be developed during the construction phase.	P	P	I	I	I	I	I	F	F									
Fin	6.3.2 Cost Recovery	H	140	**Develop a Financial Sustainability Plan** including income diversification.	P	P	P	P	I	I	I	I	I	I	F	F	F	F	O	O		
Fin	6.3.2 Cost Recovery	H	141	**Do not charge academic researchers, nor users from member and observer states for commissioned services.** Most ERICs charge users from other states and users from industry to access their services.																		
Fin	6.3.3 EU and National Contributions	H	145	**Produce an annual report: linking KPIs** (key performance indicators) **and annual budget.**	P	P	P	P	I	I	I	I	I	I	I	I	I	I	I	I	I	I
Fin	6.3.3 EU and National Contributions	H	146	**Keep an eye on EU funding calls / opportunities.**	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
Fin	6.3.3 EU and National Contributions	H	147	**Engage new Member States** to increase annual national contributions and DiSSCo fixed funding.	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
Fin	6.3.4 Pre-Commercial Procurement (PCP)	H	149	**Update the DiSSCo Construct development needs (H, prep needed)**: By listing the development needs according to their industry TRLs and their fit with the 4 strategic pillars defined during DiSSCo Prepare, we can create a list of eligible activities and their priorities, that can then serve as pipeline for PCP and PPI opportunities.								P	P	I	I	I	F	F	F	F		

				Description	1	2	3	4	5	6	7	8	9	10	11	12
				a first assessment and selection has been completed, this should be repeated at set intervals (e.g. quarterly), when intermediate perspective is compiled centrally.												
Fin	6.3.4 Pre-Commercial Procurement (PCP)	H	150	**Establish DiSSCo as a legal entity (H, prep needed)**: In order to leverage the benefits of PCP and PPI during the Construct Phase of DiSSCo, we need a legal entity or one of the consortium partners to assume the procurement role during the construction phase. At which time, also the procurement structure in DiSSCo Hub should be locked.									P	I	F	F
Fin	6.3.4 Pre-Commercial Procurement (PCP)	H	151	**Lock on DiSSCo procurement structure (H, prep needed)**: the procurement function can be centralised, decentralised, hybrid,.. Each setup has its own merits and drawbacks, nevertheless, a timely choice would benefit the future efforts in DiSSCo.									P	I	F	F
Fin	6.3.5 Impact Indicators	H	154	**Identify detailed data requirements** - Liaise with technical development and partners to provide detailed data requirements on monitoring services, tools, activities to ensure implementation of data gathering and data flow.		I	I	I	I	I	I					
Fin	6.3.5 Impact Indicators	H	155	**Establish a pilot study of indicator quantification** to establish background status of SEI indicators before operation.				I	I	I	I					
HR	7.3.1 Human Resources Policy	H	158	**Hire a HR manager to implement human resources management actions based on the policy defined.**	P	P	P	I	I	I	I	F				
HR	7.3.1 Human Resources Policy	H	159	**Identify the number and the profiles of professionals needed for DiSSCo implementation.**	P	P	P	I	I	I	I	F				

					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SysArch	4.3.4 Integration in the Global Technical Landscape	H-M	096	**Employ the developed FAIR ecosystem in the Bio*, Life and Earth System Sciences to enable advanced modelling, simulation and prediction capabilities.** This implies extending the focus from DiSSCo's core activity of NSC data management with the key feature of distributed curation during the lifecycle of the DS towards greater activities to support as well (derived) modelling lifecycles. Involved is the creation of new FDO types e.g. for machine-observations, -predictions and biologging plus an even deeper integration of tools and standards for portable workflows like the Common Workflow Language (CWL[39a]) and Research Object Crate (RO-Crate[40a]). This will build upon experiences from the Specimen Data Refinery prototyping in SYNTHESYS+ (Hardisty 2022[18b]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	M	008	**Enable machine-actionable data management for DiSSCo by providing a FAIR ecosystem of DMP-related services** Data cannot be FAIR when an infrastructure does not implement policies, rules and procedures for FAIR. Accordingly, key services to support maDMPs have to be established involving automatic recommender services for licensing, tools to compose access and update policies using machine-actionable policy elements based on machine-actionable formats like PROV-O and ODRL, and services to enable the automatic forwarding of RDM tasks to curators and other agents when specific input is expected. This involves an assessment and check for suitability of available services like the EUDAT license tool[41a].	P	P	P	P	I	I	I	I	F	F	F	F				

				Description																	
DIK	3.3.2 Data Management Plan Implementation (ICEDIG)	M	009	**Implement prospective provenance for DES** Provide a provenance model to capture the sequence of operations leading to a certain state of a DES and related biodiversity FDOs in the form of a "recipe" to facilitate restartability and partial updates from a selectable checkpoint. This should also enable the transfer of operation sequences and workflow parameters to data objects with similar constraints, e.g. the sharing of digitisation pipelines between similar subtypes of the DES. In DPP, a first prototype was developed to support "restartable DES" making use of JSON Patch[42a] and PROV-O which should be elaborated to support full workflow software environment preservation involving tools like CWLProv (Khan 2019[43a]).	P	P	P	P	I	I	I	I	F	F	F	F				
DIK	3.3.3 Digital Specimen Data Model	M	012	**Coordinate a domain expert group for peer review of modelling decisions.**	P	P	I	I			I	I	I	I	F	F	F	F	F	F
DIK	3.3.4 Semantic Annotation	M	015	During standards development, **push to support enrichment with multiple PIDs from different resources for a single concept.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
DIK	3.3.4 Semantic Annotation	M	016	**Organise and promote manual enrichment activities at individual collections** as a part of capacity enhancement.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
DIK	3.3.4 Semantic Annotation	M	017	During standards development, **push to support verbatim or raw data, so no loss of information occurs** as a result of enrichment (M).	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
DIK	3.3.5 Data Mobilisation (Institutional)	M	018	**Establish a formal but light-touch processes for updating the** <u>DiSSCo Digitisation Guides</u>[44a] **website**, reviewing and accepting new content (including workflows) to the site.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
DIK	3.3.5 Data Mobilisation (Institutional)	M	020	In collaboration with the Helpdesk, **provide a range of support channels to answer questions**, provide in-person and online <u>training</u>, and establish digitisation best practices across the DiSSCo community.	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O
DIK	3.3.5 Data Mobilisation (Institutional)	M	021	**Assess user requirements of the guides website**, including UI improvements.	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O
DIK	3.3.6 Taxonomic Backbone	M	025	**Engage further with GBIF and COL in the context of the Alliance for Biodiversity Knowledge** *Continue to support the international effort by GBIF and COL to build one common infrastructure for taxonomic names through ChecklistBank. *Engage with GBIF and COL to support the implementation of PIDs for taxonomic names and concepts.	P	P	P	P	I	I	I	I	F							
DIK	3.3.7 Knowledgebase	M	031	**Develop** <u>training</u> **and training materials** for users of the DiSSCo KB.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
DIK	3.3.7 Knowledgebase	M	032	**Further, closely link or integrate DiSSCo** <u>Helpdesk</u> (e. g. JitBit)			P	P	P	I	I	I	I	I	F	F	O	O	O	O
DIK	3.3.8 Helpdesk	M	041	**The Helpdesk should be multilingual,** offering the researcher the possibility to interact with the system using native language statements/commands. Multilingualism should be fully implemented in FAQs and	P	P	P	P	I	I	I	I	F	F						

					T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
				in support documentation, and although English will be the main working language of the Human Intermediated Q&A Service, it should have the possibility to receive questions and provide answers in different languages. Automatic detection and translation of languages should be considered.																
DIK	3.3.8 Helpdesk	M	042	**Develop supporting documentation.** Most DiSSCo services are still under development and, therefore, production of supporting documentation should be considered an ongoing process.	O	O	O	O	O	O	O	O	O	O	O	O				
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	M	048	**Situate DiSSCo as the major driver for the transformation of NSCs and other biodiversity data providers into integrated hubs for geo- and biodiversity data as part of European and global hyper infrastructures (EOSC, Gaia-X, GOSC) leveraging concepts like the FAIR Digital Twins** (Schultes 2022[45a]). This implies that the current technical focus of DiSSCo be widened by transferring and extending the established methods from collection-based research to other fields of geo- and biodiversity research (e.g. macroecology, biologging). Specifically, this means active provision and support of a FAIR ecosystem of services for the geo*- and bio*-communities involving infrastructures for modelling of new data types (beyond the current focus of the DS), hosting of registries for profiles and operations of those types and provision of service infrastructure for fundamental requirements like PIDs and AAI (cmp. section 5.3. in Wittenburg 2022[46a] for the high-level of FDO-compliance achieved by DiSSCo services).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

					P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.2 Interoperability (Institutional CMSs)	M	055	**Develop training and training materials** for the local CMS technical team.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.2 Interoperability (Institutional CMSs)	M	056	**Further, foster the creation of an ecosystem of interoperable open-source CMS modules and services across the CMS community beyond DiSSCo.** Based on the structures and tools developed within the framework of DiSSCo, the wider community of open CMS developers will be addressed to adopt support for DiSCN-linked interface standards (see D6.1 above) and data models with special emphasis on Digital Extended Specimens and various DES subtypes proposed for openDS (Addink 2022[37b]) as well as for DES/Digitisation-linked standards with special emphasis on MIDS[47a] (MIDS level 3 is closely aligned with openDS/DES).	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.3 Technical Services Development and Interoperability with DS Arch	M	060	**Establish service level agreements with identified service providers**	P	P	P	I	I	I	I	I	I	F	F	F				
SysArch	4.3.3 Technical Services Development and Interoperability with DS Arch	M	061	**Establish FDO infrastructure compliance requirements for DiSSCo services**	P	P	P	I	I	I	I	I	I	F	F	F				

Category	Ref	M	ID	Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SysArch	4.3.3.a Digital Specimen Repository (DSR) Infrastructure	M	064	**Align DSR with FDO specifications.**		P	P	P,I	P,I	P,I	F	F								
SysArch	4.3.3.a Digital Specimen Repository (DSR) Infrastructure	M	065	**Integrate DSR with the DOI ecosystem** (see PID service) Tracked in DD-5[22b] (this has dependency on a few global partners).	P	P	P,I	P,I												
SysArch	4.3.3.c Persistent Identifier (PID) Infrastructure	M	069	**Ensure FDO compliance** by including Type Registry, FDO Record, and FDO Profile. Tracked in DD-10[48a] (this is also relevant for Digital Specimen Repo[25b]).		P	P	P	P,I	P,I	P,I	P,I	F							
SysArch	4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	071	**Implement trust model**[49a] (related to UCAS enhancement work tracked in DD-26[27b]).	P	P	P,I	P,I	F	F										
SysArch	4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	072	**Seek guidance and help from EOSC and EOSC Life** expert resources via GRNet **to build a robust AAI infrastructure.**									P	P,I	P,I,F					
SysArch	4.3.3.d Authentication	M	073	**Deploy and use AAI system** for all different DiSSCo services.										P	P	P	P,I	F	F	

Team	Task	M	No.	Description								
	and Authorization Infrastructure (AAI)											
SysArch	4.3.3.f_Unified Curation and Annotation System	M	078	**Align with openDS data modelling work** for the annotation data model.	P	P	I	F				
SysArch	4.3.3.f_Unified Curation and Annotation System	M	079	**Align with FDO specification.**	P	P	P	I	I	F	F	
SysArch	4.3.3.f_Unified Curation and Annotation System	M	080	**Implement automation and machine annotation to speed up the annotation process** (needs some interaction with the SDR and other AI/ML development activities in the community).		P	P	I	I	F		
SysArch	4.3.3.g_Open Refine Data Management Tool	M	081	**Conduct surveys and consultation with users from organisations that do not have CMS.**		P	P,I	P,I	F			
SysArch	4.3.3.g_Open Refine Data Management Tool	M	082	**Provide a feedback plan to feed further development work.**				P	P	P,I	F	
SysArch	4.3.3.h_European Loans and Visits System (ELViS)	M	085	**Align ELViS with TDWG standard** (Latimer Core[32b], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information	P	I	I	F	F			

Team	Reference	M	No.	Description									
				at the collection description level). Tracked via DD-119[33b].									
SysArch	4.3.3.h European Loans and Visits System (ELViS))	M	086	**Ensure coordination and collaboration of distributed DiSSCo core architecture and ELViS code development (potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and ELViS core code base separately. Each of these teams will have different working cultures and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address incompatible workflows. It is also important to identify different milestones aligned with the feature requirements and SRL. Working in such a distributed manner might require training and workshops. The use of agile product roadmap[50a] methodology is highly recommended.	P	P	I	I	I	F			
SysArch	4.3.3.i Specimen Data Refinery (SDR)	M	088	**Integrate SDR with RO-Crate, Galaxy and FDO specification** (alignment with the BioDT[51a] and BGE[52a] project as RO-Crate might be used there as well).			P	P	P,I	,F	F		
SysArch	4.3.3.i Specimen Data Refinery (SDR)	M	089	**Provide training for local distributed technical teams in distributed ways of working (Potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and SDR core code base separately. Each of these teams will have different working culture and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps,	P	P	P	I	I	F			

				Sprint/Scrum methods can be a solution to address this problem. Working in such a distributed manner might require training and workshops.																
SysArch	4.3.3.j Collections Digitisation Dashboard (CDD)	M	092	**Align CDD with the DiSSCo core infrastructure development** activities. Related to DiSSCo DD-16[30c].				P	P	P	I	I	I	F	F	F				
SysArch	4.3.4 Integration in the Global Technical Landscape	M	097	**Assert DiSSCo's role as the major driver of the converging and harmonisation process of the Digital Extended Specimen with respect to establishing a joint digital object approach. Ensure compatibility of data models within approaches** aiming to provide digital object-based information entities like DiSSCo's Digital Specimen and BCoN's Extended Specimen with wider _interoperability_ standards. It is significant that in particular outcomes of the Digital Extended Specimen conversion discussions (Webster 2021[53a]) comply with edge conditions set by overarching FDOt standards (e.g. on FDO Typing).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.5 Geo-collections Mobilisation	M	101	**Explore a potential relationship with Mindat for automated integration between DiSSCo and mindat.org[54a].** • Collaborate with Mindat to develop a specification and resourcing estimate for integration using DiSSCo's event-driven approach		P	I	I	I											

Theme	Ref		#	Task	Timeline
				• Investigate potential use cases and appetite for the integration or linkage of DiSSCo Digital Specimen data in the Mindat platform	
SysArch	4.3.5 Geo-collections Mobilisation	M	102	**Work closely with GeoCASe on an early proof of concept for event-driven interoperability with an external service** Use the process to: • develop a blueprint and set of minimum requirements for external services to interoperate with DiSSCo using the event-driven architecture; • develop a detailed specification, resource requirements and costings for GeoCASe development to automate ingestion from DiSSCo using the event driven approach.	P I I F
GSS	5.3.2 Governance Strategy and Strategic Implementation	M	114	**Keep principles of transparency, flexibility and agile decision-making process and Design thinking methodology**, as pillars for updating the Strategy.	O O O O O O O O O O O O
Fin	6.3.1 Cost Book	M	139	**Organise a meeting with national nodes to share / review the area concepts.** Can be developed during the construction phase.	P P I I I I I F F
Fin	6.3.2 Cost Recovery	M	142	**Diversify sources of income for sustainability of the ERIC** by: (a) creating interactions and partnerships between the ERIC and different stakeholders, both in the public and private sectors, at several levels (International, European, national and regional), and (b) diversifying the services provided by DiSSCo-ERIC and apply a charge to access them.	

				Task	Timeline
Fin	6.3.2 Cost Recovery	M	143	**Develop a Central Business Development team to further develop a cost recovery framework** including trialing some of the proposed business models, and managing a portfolio of income diversification projects and services.	
Fin	6.3.2 Cost Recovery	M	144	**DiSSCo ERIC should not be constrained in charging for services in its statutes** when specifying a maximum level of income from economic activities.	
Fin	6.3.3 EU and National Contributions	M	148	**Provide** training **regarding submitting funding applications and consortium building**	P P P P I I I I I I I I
Fin	6.3.4 Pre-Commercial Procurement (PCP)	M	152	**Issue a Prior Information Notice (PIN) (M, provided the pre-commercial procurement high priority recommendations have been completed)**: with a refined scope of the proposed development, performing a gap analysis to track existing solutions and their TRL levels to define the preferred procurement strategy (PCP, PPI, standard procurement),	P P I I F F
Fin	6.3.4 Pre-Commercial Procurement (PCP)	M	153	**Work out the PCP and PPI contract notices (M, provided the pre-commercial procurement high priority recommendations have been completed)**: Engage in the preparatory work for organising PCP and PPI calls, which includes a clear scope definition, preliminary market consultation leading up to a publication of the contract notice, which details the problem to be solved, its functional requirements and use cases.	P P I I F F
Fin	6.3.5 Impact Indicators	M	156	**Define roles within the DISSCo organisation about data gathering, analysis and reporting of SEI indicators**	

				Task																				
Fin	6.3.5 Impact Indicators	M	157	**Start the design and implementation of the information management system for data gathering, analysis, reporting and archiving**. It might require new developments to implement automatisms;																		
HR	7.3.1 Human Resources Policy	M	160	**Define a compensation plan that can be attractive to the selected professional.** The figures will be necessary for DiSSCo implementation.	P	P	P	P	I	I	I	I	I	I	I	F				
HR	7.3.1 Human Resources Policy	M	161	**Define a recruitment strategy and design a professional training program to acquire the skills required by DiSSCo implementation.** Training should go beyond job-related skills development like project management, IT development, open science, law, EU projects or service delivery, and include skills focusing on long-term personal development such as leadership, communication, networking skills, stakeholder engagement, project management, presentation skills, facilitation skills, IT and informatics skills and so on.	P	P	P	P	I	I	I	I	I	I	I	F				
HR	7.3.1 Human Resources Policy	M	162	**Write specific guidelines for mobility, work conditions and equality of treatment.**	P	P	P	P	I	I	I	I	I	I	I	F				
HR	7.3.3 Capacity Self-Assessment (Digital Skills including Dashboard)	M	171	**Ensure software development for digital maturity self-assessment tool:** The digital maturity self-assessment tool is being developed under T7.3 but will require ongoing support and refinement including digital maturity content. Please refer to T7.3 for recommendations on future development needs for this tool. (M)																
HR	7.3.3 Capacity Self-Assessment	M	172	**Define questions for sub-categories of digital maturity and self-assessment tools:** MS3.4 details the high level																

					T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
	(Digital Skills including Dashboard)			content and subcategories for the digital maturity tool, and examples of content for the self-assessment, but further work will be required to define questions in each sub-category. This should link closely with outputs from WP2, to ensure the self-assessment tool can support institutions and individuals by linking to available <u>training</u> and resources. (M)																
HR	7.3.3 Capacity Self-Assessment (Digital Skills including Dashboard)	M	173	**Ensure role/job profiles are setup in the <u>Knowledgebase</u> (KB):** This task recommends that relevant role/job profiles are set up in the DiSSCo KB. This work is being progressed in WP2.																
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch))	M-L	049	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation [H] (concept: "<u>UCAS - Unified Curation and Annotation System</u>") and **annotation [M-L] (**prototypes Specimen Data Refinery <u>Hardisty 2022[18c]</u>, Machine Learning as a Service (MLaaS) pilot Grieb 2021).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
DIK	3.3.6 Taxonomic Backbone	L	026	**Engage with the TDWG Paleo Data Working Group on representing fossil taxonomy in the <u>Digital Specimen Data Model</u> and consult with COL on improvements in ChecklistBank**							P	I	I	F						
DIK	3.3.7 Knowledgebase	L	033	**Further, integrate DiSSCo Training software** (e. g. Moodle)		P	P	P	I	I	I	I	I	F	F	O	O	O	O	
DIK	3.3.7 Knowledgebase	L	034	**Further, research and integrate further functionality upon request.**		P	P	P	I	I	I	I	I	F	F	O	O	O	O	

				Task																
SysArch	4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	L	050	**Provide a discovery service to cluster/match types of Digital Specimens (and other biodiversity Digital Objects) against requirements for scientific data**. Objective is to enable the determination of semantic similarity of DS and the suitability of the encapsulated data by providing (or ranking) of appropriate Digital Objects using DS attributes transformed to vector representations/embeddings (Le 2014[55a], Devlin 2019[56a], Wittenburg 2022 in sub.)	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
SysArch	4.3.2 Interoperability (Institutional CMSs)	L	057	**Further, evaluate, develop and integrate further functionality** upon request.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
SysArch	4.3.5 Geo-collections Mobilisation	L	103	**Implement the changes specified in the GeoCASe early proof of concept** if appropriate resources can be made available.									P	I	I	F				
GSS	5.3.2 Governance Strategy and Strategic Implementation	L	115	**Ensure DiSSCo Strategy implementation is updated regularly** during the transition and construction phases;	O	O	O	O	O	O	O	O	O	O	O	O	O			
GSS	5.3.5 National Nodes Engagement	L	126	**(L) Create internships and/or student grants at institutional level to overcome detected difficulties (in terms of resources availability)** for collecting relevant information from the different members of each node. This action can be coordinated through CETAF.			P	P	I	I	I	I	F							
HR	7.3.1 Human Resources Policy	L	163	**Design a training program on multidisciplinary teamwork to enhance the effectiveness of teams.**	P	P	P	P	I	I	I	I	I	I	I	F				

					Timeline																			
HR	7.3.1 Human Resources Policy	L	164	With the purpose of ensuring compliance with high quality standards in the exercise of DiSSCo's functions **it is relevant to carry out systematic assessments/audits** in order to identify and correct non-conformities, as well as the creation of dedicated institutional communication channels in order to allow the dissemination of clear, concise and useful information about DiSSCo.	P	P	P	P	I	I	I	I	I	I	I	F								
GSS	5.3.4 DiSSCo Policies and Procedures		119	**Further develop the metadata schema of the policy tool to cover the full range of DiSSCo services**, once policy requirements for these services have been determined. (The metadata schema developed in SYNTHESYS+ NA2.1 only covers the policy requirements for ELViS. Further work will be required.) The metadata schema and policy tool are designed to be extensible, and further policy components can be added.																				
GSS	5.3.4 DiSSCo Policies and Procedures		120	**Improve the institutional summaries and DiSSCo CSO dashboard.** It is hoped that a proof-of-concept summary and dashboard will be created as part of 7.3, using ELViS as an example, however, more development will be required once policy requirements for other DiSSCo Services have been agreed.																				
GSS	5.3.6 National Smart Specialisation Plan		127	The Specialisation tool is delivered as a "tested" tool on a sample of institutions from the NH community and Nodes. Although all the NN currently part of DiSSCo have been contacted the current data entered only represent a limited number of data and thus it will require additional efforts to **obtain from all NNs a representative set of data** leading to the analysis of the specialisations present throughout the community of NH institutions.	I	I	I	I	F	F	F													

GSS	5.3.6 National Smart Specialisation Plan		128	**Ensure at least two institutions per node provide data.** The Specialisation tool gives access to the characteristics and assets of research institutions and nodes that are part of the community of natural history institutions. It is therefore recommended that at least 2 institutions per node provide their data to be able to have a critical mass of data available to search into and enable the expected analysis of key specialisations. The development of the tool is finished, although it can further evolve and be refined. Its implementation "I" phase is ongoing with enlarged data sets provided from the NH institutions, with a minimum of 2 institutions per NN having contributed to the data set. We consider that an additional year would be needed to reach this level. The Finalisation phase will be reached when the analysis will be applicable to a representative set of institutions of at least 2 institutions per Node.	I	I	I	I	F	F	F												
GSS	5.3.6 National Smart Specialisation Plan		129	**Manage and maintain the Specialisation tool** by a representative of the network and community of natural sciences institutions. The information held in the database represents key assets that should be used diligently. This information will also be enriched by further users of the DiSSCo RI that will rely on the FAIR access and GDPR protocols which will ensure a trustworthy repository.	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
GSS	5.3.7 Communication and Engagement		130	**Revise and adapt the Communication Strategy according to the next phases of DiSSCo.** Communication work is a non-stop task that necessarily will evolve as DiSSCo develops, and needs to be customised according to the RI's evolution. Therefore, actions in communication		P	P	I	F													

				T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
			need to be maintained and furthermore, increased during the construction phase towards the operation launch of DiSSCo.																
GSS	5.3.7 Communication and Engagement	131	**Define and implement an evolving schema for communication material** so that users and interested agents can easily understand the status-quo of DiSSCo, the progress made, and the outstanding stages to address;		P	P	I	I	I	I	F								
GSS	5.3.7 Communication and Engagement	132	**During Transition further develop and put into practice, the figure of DiSSCo Ambassador,** outlined during the preparatory phase to address strategically the constitution of the DiSSCo ERIC and later on, during the construction phase, the expansion of its membership;		P	P	I	I	I	I	I	I	I	I	F				
GSS	5.3.7 Communication and Engagement	133	**Encourage increased alignment with sister initiatives in Europe and globally to continue the ongoing processes of linking RIs acting in the field,** through either joint projects such as BiCIKL, or collaborative strategic partnerships.			P	I	I	I	I	I	I	I	I	I	I	I	I	F
GSS	5.3.7 Communication and Engagement	134	**Maintain a person fully dedicated to DiSSCo communication tasks** to keep the targeted audiences duly updated and well informed, and to embed these activities also into the development of the DiSSCo ERIC.	P	P	I	I	I	I	F									
GSS	5.3.8 Strategic Partners Engagement	135	**Procurement Strategy:** The **MS8.4 milestone** recommends that DiSSCo develops a procurement strategy in order to align long term priorities and objectives with procurement processes. **This strategy should help DiSSCo to develop partnerships, scale up**																

				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
			processes, as well as mitigate risks and improve cost efficiency. This work is continued by T4.4. Stakeholder Analysis: MS8.6 details DiSSCo stakeholders, including GBIF, Catalogue of Life, GeoCASe, TDWG and International Barcode of Life. This report informs stakeholder engagement plans and prioritisation.															
GSS	5.3.9 Advocacy and Engagement	136	Revise and adapt the Advocacy and Engagement Strategy according to the next phases of DiSSCo.	P	P	I	I	I	I	I	I	I	I	I	I				
HR	7.3.2 Training Strategy	168	During the construction phase and towards practical implementation of training, other activities shall be supported: Integrate DiSSCo online training into the e-learning platform provided by TETTRIs project.		P	P	I	I	I	I	I	I	I	I	F				
HR	7.3.2 Training Strategy	169	During the construction phase and towards practical implementation of training, other activities shall be supported: Link training provision with expertise availability by creating a comprehensive lively web-based database of experts in different disciplines and fields necessary for digitisation work.					P	P	P	I	I	I	I	I	F			
HR	7.3.2 Training Strategy	170	During the construction phase and towards practical implementation of training, other activities shall be supported: Investigate recognition support with academia.					P	P	P	P	I	I	I	F				
HR	7.3.4 Distributed Teams and Secondment	174	Continue to evaluate distributed ways of working throughout construction. The distributed working recommendations will be relevant throughout the construction phase of DiSSCo.																

| HR | 7.3.4 Distributed Teams and Secondment | 175 | **Continue to evaluate secondment (tech infrastructure developers) to CSO during construction.** The secondment recommendations contained within MS3.9 may be relevant during the construct phase of DiSSCo, depending on whether secondment to the DiSSCo Coordination and Support Office (CSO) is seen as beneficial during this period (e.g. for development of the technical infrastructure). However, distributed team-working alongside existing roles may be a more appropriate model. There are no current construction tasks as no decision has been taken to implement a secondment model. | | | | | | | | | | | | | | | |

03 Data, Information and Knowledge

3.1 Overview
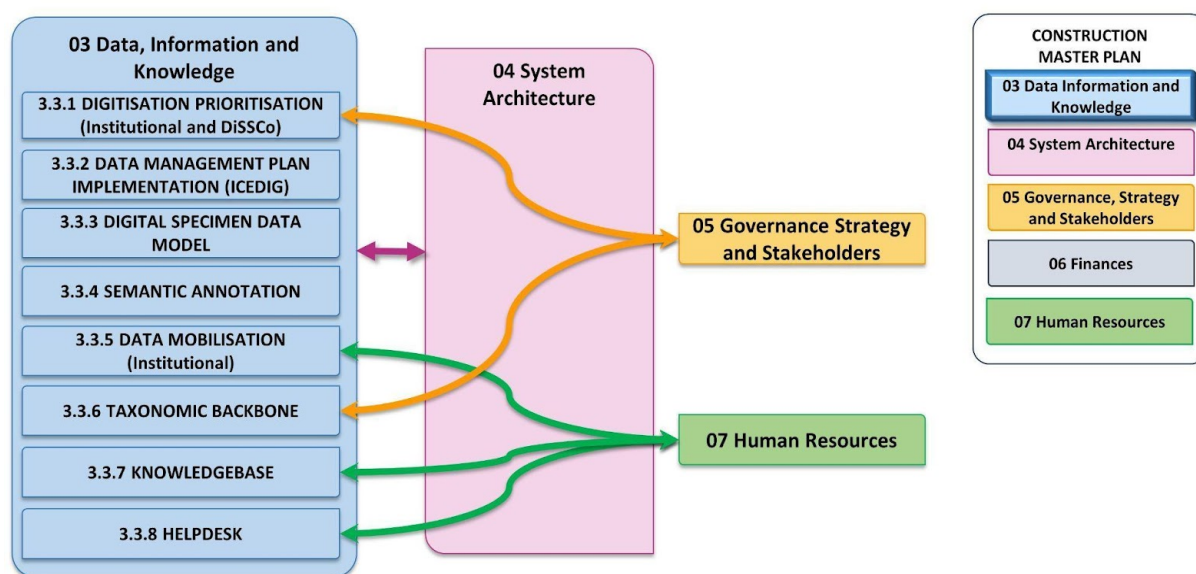
Figure 2. Synoptic graph showing the dependencies (arrows) between tasks in this chapter with other tasks or chapters (systems) in DiSSCo Construct.

Generating, curating, publishing, and using data are key functionalities for all of the DiSSCo e-services. This section provides an overview of the key elements in the DiSSCo FAIR data lifecycle. These elements are the cornerstone for the data, information, and knowledge that will be utilised by the system architecture. The major outcome of these tasks are understanding and describing the nature of the data in the DiSSCo universe along with creating the necessary support for data mobilisation and knowledge dissemination.

Digitisation workflows are essential steps for creating Digital Specimens. Since there are billions of specimens to digitise, digitisation coordination among DiSSCo partners will ensure prioritisation. Prioritisation activities need to follow guidelines which will ensure relevance and data quality. At the same time prioritisation needs to be aligned with scoring and review processes. A Data Management Plan (DMP) therefore will play a key role in all of these activities along with data modelling and specification work. In particular the Digital Specimen Data Model, part of the open Digital Specimen (openDS) specification focuses on

what Digital Specimens should look like and MIDS (Minimum Information about a Digital Specimen) establishes the minimum data elements needed for Digital Specimens. All the digital objects within DiSSCo need to adhere to the FAIR principles (Findable, Accessible, Interoperable and Reusable) where linking, integration and interoperability are different. Therefore, implementing FAIR principles via the FAIR Digital Object framework is also relevant for data generation and management.

This chapter provides:

- An overview of the work done in the DiSSCo preparation phase to support the exchange of data, information and knowledge including the creation of services for knowledge sharing and support (Helpdesk and Knowledgebase);

- An overview of the essential characteristics for data management of Digital Extended Specimen;

- Timeline and summary recommendations for Data Management Plan Implementation and Digitisation Prioritisation;

- Recommendations how to further develop the Digital Specimen Data Model and Semantic Annotation support;

- A description of the actions needed to support a global Taxonomic Backbone by engaging with GBIF (Global Biodiversity Information Facility), COL (Catalogue of Life) and the TDWG (Biodiversity Information Standards, formerly Taxonomic Databases Working Group) Paleo Data Working Group.

3.2 Timeline and Summary Recommendations

Table 2. Data, Information and Knowledge Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Task Name	Pri Typ	Rec No	Recommendation	2023				2024				2025				2026			
			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
				1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3.3.2 Data Management Plan Implementation (ICEDIG)	GAP	010	"Machine-actionability", the capability of machines to handle data autonomously and appropriately is a core objective of FAIR and thus FDOs (Jacobsen 2020[7b]). The concept was i.a. developed to describe an aspired transformation of DMPs from essentially static documents towards an application-based tool chain integrating and automatising all relevant information steps in the data life cycle: the machine-actionable data management plan (maDMP; Miksa 2019[8b]). **Due to several constraints in DPP, a merge of the maDMP concepts and the FDO approach wasn't elaborated to the same extent as a prototype in DPP like the machine-learning and human- (crowd-) based annotation pipelines** (cp. Leeflang 2022[9b]; Grieb 2021[10c]).																

			The objective and potential of the maDMP are the seamless integration of the reporting tools into the existing network of project tools and services to release involved humans (developers, scientists, curators etc) from the "annoying administrative exercise" (Miksa 2019) to write a DMP as a matter of form. It should be feasible to realise this level of automatisation for DiSSCo's Digital Specimen services, since the fundamental Digital Object Architecture provides already a network of FAIR-enabling services, which (prototypically) implement the policies, rules, procedures and infrastructures for specimen data during its whole life cycle. **The main task now is to provide a proper alignment on several levels (data model, services) with common specifications (DCAT, PROV-O) to make the DES more "speech-enabled" with regard to retrospective and prospective provenance.**																					
3.3.2 Data Management Plan Implementation (ICEDIG)	Tr	004	**Further develop and refine the provenance model in openDS**.	I	I	I	I																	
3.3.2 Data Management Plan Implementation (ICEDIG)	Tr	005	**Provide a more complete capture of digitisation processes in MIDS** by establishing a DES as output of a MIDS level 3 description (align MIDS information elements with openDS) with the aim to facilitate the transfer of digitisation workflows.	I	I	I	I																	
3.3.1 Digitisation Prioritisation	H	001	**We recommend collaboration on digitisation projects,** especially at DiSSCo level, but the criteria for prioritisation are also applicable to institution-specific,	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

(Institutional and DiSSCo)			national or regional projects. Coordination among DiSSCo partners will support the community itself and the rapid developments in approaches which are happening around the world as a solution to help drive forward strategic prioritisation of digitisation activities. Communicating summaries of these and adding to these will have a dual role in helping others define or refine their strategies.																
3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)	H	002	**We suggest that DiSSCo and its partners apply the guidelines for prioritisation of digitisation developed by Task 1.3,** which suggest to **focus on the four main criteria established, i.e. Relevance, Data quality, Cost and Feasibility, and to answer the relevant questions that were provided.**	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)	H	003	**For the evaluation process we recommend a combination of a scoring system and a panel review,** similar to how applications have been evaluated in the SYNTHESYS projects[14b]. We consider a panel review to be necessary because of the diversity and multi-dimensionality of the prioritisation criteria. We are aware that the guidelines may need modification according to future experience gained.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
3.3.2 Data Management Plan Implementation (ICEDIG)	H	006	**Establish a FDO type for a machine-actionable DMP** Link all relevant outputs of a DES' lifecycle to a maDMP-FDO comprising links to retrospective and where applicable prospective provenance objects, to related services like re3data for repository information, to machine-actionable policies in a structured format	P	P	P	P	I	I	I	I	F	F	F	F				

			like ODRL. Elements of this FDO type must support a sufficient level of granularity (no long free text as bitstream) to enable reaggregation of information elements to adjust the DMP dynamically.														
3.3.2 Data Management Plan Implementation (ICEDIG)	H	007	**Implement linking to comparable service networks like ENVRI-FAIR** Enable the interoperability of DiSSCo's core entities involved in RDM by providing mappings to a common vocabulary shared among relevant data providers like the ENVRI-FAIR network. A common standard here is W3C DCAT, which is centered on requirements of data exchange of data portals in Europe providing semantic interoperability by reuse of (e.g. EuroVoc) and mappings to existing vocabularies (e.g. INSPIRE, schema.org[15b]). Predefined profiles are available like DCAT-AP used for EU infrastructures, EPOS DCAT-AP employed in ENVRI-FAIR which provides starting points for DiSSCO's mapping (or subclassing) of core entities like DES, collection, CMS, enrichment service to dcat:Resource, dcat:catalog dcat:Service etc.	P	P	P	P	I	I	I	I	F	F	F	F			
3.3.3 Digital Specimen Data Model	H	011	We foresee that the basic modelling processes will have been established with the end of the DiSSCo preparation phase. However, there will be the need for further coordination of works on the model itself. The DiSSCo synchronisation group for data standards strongly recommended allocating resources for the coordination of modelling activities. **Allocate resources for the coordination of further modelling activities and recruit a modelling expert.**	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F	F

Section	Priority	No.	Recommendation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
3.3.4 Semantic Annotation	H	013	**Develop tools that enable automated enrichment** both as part of the digitisation process and as part of publication pipelines.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
3.3.4 Semantic Annotation	H	014	**Organise a working group to exchange experiences on enrichment, and discuss and promote curation of open resources** such as Wikidata to avoid double-work.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
3.3.5 Data Mobilisation (Institutional)	H	019	**Establish a Digitisation Interest Group** with representation and linkage to other relevant community groups (e.g., TDWG, CETAF DWG/ISTC) to continue development of the DiSSCo Digitisation Guides, and discuss challenges and solutions for all aspects of digitisation.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
3.3.5 Data Mobilisation (Institutional)	H	022	**Determine how much, if any, investment DiSSCo plans to make in digitisation infrastructure and how much support will be provided to institutions to increase their digitisation capacity.** This includes support to keep the Digitisation Guides up-to-date and to foster community curation. This would ideally be combined with both in-person and online <u>training</u>, and potentially support planning and implementation of physical and digital infrastructure investment. We recommend focusing content and support on new or recently established national nodes where digitisation is a priority, and national nodes which are working on national or regional digitisation programmes.	P	P	P	P	I	I	F	F	O	O	O	O	O	O	O	O
3.3.6 Taxonomic Backbone	H	023	**Focus on articulating more detailed requirements for taxonomic names services**		P	I	F												

3.3.x Section		No.	Description	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
			*Species 2000 would recommend a further in-depth exploration with DiSSCo into the requirements for taxonomic names services. *Engage with COL to further explore the benefits that COL and ChecklistBank might leverage from interoperability with DiSSCo, and resource implications of supporting those requirements in the DiSSCo architecture.	P														
3.3.6 Taxonomic Backbone	H	024	**Assess and refine the potential technical approaches to interoperability with ChecklistBank** *Use technical pilots to assess the possibilities and limitations of using COL's public services and open APIs to meet DiSSCo integration requirements. *Engage with COL to explore the potential implementation of the event-based interoperability model, and any associated investment that might be required into COL development.	P	I	I	I	F										
3.3.7 Knowledgebase	H	027	**Establish a dedicated person (editor) or editorial team** responsible for quality control, content management and curation of the DiSSCo KB.		P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.7 Knowledgebase	H	028	**Engage or provide technical maintenance services and basic software support** for KB.		P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.7 Knowledgebase	H	029	**Further, integrate with DiSSCo's Authentication and Authorization Infrastructure** (AAI).		P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.7 Knowledgebase	H	030	**Further, implement permanent ongoing curation, maintenance and development** (based on future requirements) of the service to remain a trusted information source.		P	P	P	I	I	I	I	I	F	F	O	O	O	O

3.3.8 Helpdesk	H	035	**Offer three complementary services for DiSSCo Helpdesk: FAQs - Frequently Asked Questions; Human Intermediated Question and Answer Service; and Supporting Documentation.** With respect to supporting documentation, most DiSSCo services are still under development and, therefore, production of supporting documentation should be considered an ongoing process.	P	P	P	P	I	I	I	I	F	F							
3.3.8 Helpdesk	H	036	**For DiSSCo Helpdesk use the JitBit Helpdesk Ticketing System adopted for ELViS helpdesk.** Considering that the DiSSCo community is already familiar with this system, we anticipate no major issues, and no further preparation activities are needed to initiate its construction in 2024. We are confident that its construction will end by 2025.	P	P	P	P	I	I	I	I	F	F							
3.3.8 Helpdesk	H	037	**Integrate DiSSCo Knowledgebase (KB) with DiSSCo Helpdesk** which will provide links to new and/or relevant content; automatically generate search results based on incoming questions, supporting personalised responses; and linking to other DiSSCo eServices, etc.	P	P	P	P	I	I	I	I	F	F							
3.3.8 Helpdesk	H	038	**Follow the recommendations of the Web Accessibility Initiative (WAI) for the Helpdesk**, embracing an inclusive design that meets the needs of those who are permanently or temporarily disabled.	P	P	P	P	I	I	I	I	F	F							
3.3.8 Helpdesk	H	039	Allow real-time monitoring of the support process in The Helpdesk, namely the Human Intermediated Q&A Service. **Develop a Service Level Agreement**	P	P	P	P	I	I	I	I	F	F							

				P	P	P	P	I	I	I	I	F	F	F	F				
			(SLA), in which goals are set to meet users' expectations.																
3.3.8 Helpdesk	H	040	**Provide training to support staff on the policies and procedures, as well as on the tools and technology** they will be using. This will help ensure that the Helpdesk team is able to provide high-quality support to users.	P	P	P	P	I	I	I	I	F	F						
3.3.2 Data Management Plan Implementation (ICEDIG)	M	008	**Enable machine-actionable data management for DiSSCo by providing a FAIR ecosystem of DMP-related services** Data cannot be FAIR when an infrastructure does not implement policies, rules and procedures for FAIR. Accordingly, key services to support maDMPs have to be established involving automatic recommender services for licensing, tools to compose access and update policies using machine-actionable policy elements based on machine-actionable formats like PROV-O and ODRL, and services to enable the automatic forwarding of RDM tasks to curators and other agents when specific input is expected. This involves an assessment and check for suitability of available services like the EUDAT license tool[41b].	P	P	P	P	I	I	I	I	F	F	F	F				
3.3.2 Data Management Plan Implementation (ICEDIG)	M	009	**Implement prospective provenance for DES** Provide a provenance model to capture the sequence of operations leading to a certain state of a DES and related biodiversity FDOs in the form of a "recipe" to facilitate restartability and partial updates from a selectable checkpoint. This should also enable the	P	P	P	P	I	I	I	I	F	F	F	F				

transfer of operation sequences and workflow parameters to data objects with similar constraints, e.g. the sharing of digitisation pipelines between similar subtypes of the DES. In DPP, a first prototype was developed to support "restartable DES" making use of JSON Patch[42b] and PROV-O which should be elaborated to support full workflow software environment preservation involving tools like CWLProv (Khan 2019[43b]).

3.3.3 Digital Specimen Data Model	M	012	**Coordinate a domain expert group for peer review of modelling decisions.**	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F	F
3.3.4 Semantic Annotationn	M	015	During standards development, **push to support enrichment with multiple PIDs from different resources for a single concept.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
3.3.4 Semantic Annotation	M	016	**Organise and promote manual enrichment activities at individual collections** as a part of capacity enhancement.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
3.3.4 Semantic Annotation	M	017	During standards development, **push to support verbatim or raw data, so no loss of information occurs** as a result of enrichment (M).	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
3.3.5 Data Mobilisation (Institutional)	M	018	**Establish a formal but light-touch process for** updating the DiSSCo Digitisation Guides[44b] website, reviewing and accepting new content (including workflows) to the site.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
3.3.5 Data Mobilisation (Institutional)	M	020	In collaboration with the Helpdesk, **provide a range of support channels to answer questions**, provide in-person and online training, and establish	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O

				C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
			digitisation best practices across the DiSSCo community.																
3.3.5 Data Mobilisation (Institutional)	M	021	**Assess user requirements of the guides website**, including UI improvements.	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O
3.3.6 Taxonomic Backbone	M	025	**Engage further with GBIF and COL in the context of the Alliance for Biodiversity Knowledge** *Continue to support the international effort by GBIF and COL to build one common infrastructure for taxonomic names through ChecklistBank. *Engage with GBIF and COL to support the implementation of PIDs for taxonomic names and concepts.	P	P	P	P	I	I	I	I	F							
3.3.7 Knowledgebase	M	031	**Develop training and training materials** for users of the DiSSCo KB.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.7 Knowledgebase	M	032	**Further, closely link or integrate DiSSCo Helpdesk** (e. g. JitBit)			P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.8 Helpdesk	M	041	**The Helpdesk should be multilingual,** offering the researcher the possibility to interact with the system using native language statements/commands. Multilingualism should be fully implemented in FAQs and in support documentation, and although English will be the main working language of the Human Intermediated Q&A Service, it should have the possibility to receive questions and provide answers in different languages. Automatic detection and translation of languages should be considered.	P	P	P	P	I	I	I	I	F	F						
3.3.8 Helpdesk	M	042	**Develop supporting documentation.** Most DiSSCo services are still under development and, therefore,	O	O	O	O	O	O	O	O	O	O	O	O				

			production of supporting documentation should be considered an ongoing process.																		
3.3.6 Taxonomic Backbone	L	026	**Engage with the TDWG Paleo Data Working Group on representing fossil taxonomy in the Digital Specimen Data Model and consult with COL on improvements in ChecklistBank**							P	I	I	F							
3.3.7 Knowledgebase	L	033	**Further, integrate DiSSCo Training software** (e. g. Moodle)			P	P	P	I	I	I	I	I	F	F	O	O	O	O
3.3.7 Knowledgebase	L	034	**Further, research and integrate further functionality upon request.**			P	P	P	I	I	I	I	I	F	F	O	O	O	O

3.3 Task Summaries

3.3.1 Digitisation Prioritisation (Institutional and DiSSCo)

Establish relevant criteria to identify a prioritisation model for digitisation

Description   *DPP D1.3*

This task relates directly to the following objective within DPP WP1: to identify criteria for establishing a priority for the digitisation of natural science collections. The outcome of the task is DPP deliverable D1.3 "Report on relevant criteria for prioritisation of the digitisation" which includes guidelines for prioritisation at DiSSCo and institutional level.

Timeline and Summary Recommendations

Table 3. Digitisation Prioritisation (Institutional and DiSSCo) Recommendations

NB:Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	001	**We recommend collaboration on digitisation projects,** especially at DiSSCo level, but the criteria for prioritisation are also applicable to institution-specific, national or regional projects. Coordination among DiSSCo partners will support the community itself and the rapid developments in approaches which are happening around the world as a solution to help drive forward strategic prioritisation of digitisation activities. Communicating summaries of these and adding to these will have a dual role in helping others define or refine their strategies.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
H	002	**We suggest that DiSSCo and its partners apply the guidelines for prioritisation of digitisation developed by Task 1.3,** which suggest to **focus on the four main criteria**	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

		established, i.e. Relevance, Data quality, Cost and Feasibility, and to answer the relevant questions that were provided.																			
H	003	**For the evaluation process we recommend a combination of a scoring system and a panel review,** similar to how applications have been evaluated in the SYNTHESYS projects[14c]. We consider a panel review to be necessary because of the diversity and multi-dimensionality of the prioritisation criteria. We are aware that the guidelines may need modification according to future experience gained.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

Discussion and Conclusions

Based on previous studies and on the analysis we carried out, the criteria for prioritisations were classified into four main categories: Relevance, Data quality, Cost and Feasibility.

All criteria that were suggested in previous studies fall into one (or more) of the four categories which are thus not new criteria but were established as an aid to reduce the multi-dimensionality of the "criterion space" during the first steps in the prioritisation process.

The outcome of our analyses of the surveys, questionnaires, and case studies indicate that, when choosing what to digitise and how to do it, it is pivotal to consider:

- Potential collaborations on digitisation proposals, particularly within the DiSSCo framework (eg. DiSSCo consortium members).

- Aiming to provide data that are sufficient for the use case within the project, whilst considering other likely use cases and paying attention to data quality.

- Using a combined approach of scoring and panel review for the implementation of the prioritisation criteria (Bakker et al. 2018 57).

More specifically, the main attributes for each category that must be considered are:

- Relevance: scientific, collection, economic and societal relevance

- Data quality: information level (e.g., MIDS (Minimum Information about a Digital Specimen) and precision), reliability (e.g. accuracy) and possibilities for data cleaning and validation)

- Cost: pre-digitisation, digitisation sensu strictu and post-digitisation costs.

- Feasibility: availability of staff skilled in collection work, staff skilled in digitisation/IT (Information Technology), IT and collection infrastructure.

Because the criteria are interdependent and complex, and because the preconditions for digitisation projects are so variable, it was deemed impossible to come up with a simple "algorithm" or "decision tree" for prioritisation.

We provided a list of questions in each of the four categories to be addressed when planning a digitisation study, or when evaluating proposals for such studies. A few examples of questions are:

- What is the scientific relevance of the project? (Which types of research will be facilitated by the generated digital data?)

- Is the cost/benefit ratio of the project reasonable? ("value for money"). Are all steps in the digitisation process considered?

- Is the level of information (e.g., MIDS) of the generated digital data sufficient for the purpose of the project?

- Is the necessary scientific (e.g., taxonomic experts, curators) and technical (e.g., IT) staff available? If not, is there a realistic plan for making such staff available?

Method

A literature search was made to supplement the very extensive analysis of criteria carried out by the ICEDIG[1b] (Innovation and consolidation for large scale digitisation of natural heritage) project.

In addition, an analysis was made of digitisation plans and criteria being used by DiSSCo partners through two surveys:

1) Essay based questionnaire: DiSSCo partners were asked to provide information, in free text, on their digitisation strategy (whether they have it) and prioritisation criteria employed for digitisation which has already been done or is in progress in the institution. A list of helpful questions to focus on the main information needed was supplied (e.g., Do you have a clear overview of the digitisation status of your institution? What is your digitisation level: specimen level or higher collection unit level? etc.). It was also suggested to distinguish between:

   - Mass digitisation or large scale where indeed the questions of prioritisation, feasibility etc. are very relevant

   - Digitisation on demand

   - Opportunistic digitisation

2) Multiple choice questionnaire: from the information gathered through the essay based questionnaire, a short multiple-choice questionnaire focused on the digitisation activity was developed using a Google Form. The short questionnaire, after being reviewed by the task partners, was sent to all DiSSCo National Nodes who shared it with their own institutions in order to collect information from as many institutions as possible involved in DiSSCo. To facilitate the dissemination, the questionnaire was translated into different languages (English, Danish, French, Italian and Dutch). The structure of the questionnaire was as follows:

· Q1 – Q3 compiler's information (personal details, e-mail, role, country, institution)

· Q4 – Q5 information about collections (size and staff employed)

· Q6 – Q9 information about digitisation strategy (digitisation initiative, digitisation priorities classified in 5 main categories, Scientific Relevance, Institutional Relevance, Economic Relevance, Educational Relevance, Technical feasibility and subcategories for each one of them)

· Q10 – Q12 information about the management of collections (overview and monitoring of the digitisation status, use of CMS - collection management system)

· Q13 – Q16 information about digitised items (procedure for data validation, standard used for databasing, digitisation levels for databased items, images and 3D models)

· Q17: further remarks about digitisation strategy

Finally, two case studies of collection digitisation provided by the Task partners were analysed in order to gather concrete information on cost/feasibility and make helpful considerations for planning digitisation activity.

3.3.2 Data Management Plan Implementation (ICEDIG)

Embedding DiSSCo in the technical landscape

Description   *DPP D6.4*

DMPs (Data Management Plan) are documents guiding actions and liabilities of research data management during the whole data lifecycle including project outputs and involving the maintenance of correct attribution and provenance data. For DiSSCo, this means that provenance data should be generated and preserved by all operations acting upon DiSSCo core data objects, the Digital Extended Specimens. A detailed set of requirements was developed under the lead of Alex Hardisty in the framework of the ICEDIG[1c] project in the form of the *Provisional Data Management Plan for the DiSSCo infrastructure* (Hardisty 2019[58]). Here the DMP is defined as a living document, i.e. a document that is continually

supplemented and updated based on the technical advancement of DiSSCo's data models and infrastructures as well as adjusted in response to developments in the global and in particular Europe's scientific and technological landscape. Accordingly, participants in WP6/T6.4 identified two developments of particular relevance: (i) The increasing differentiation of DiSSCo's core data model, the DES (Digital Extended Specimen) (Islam 2020[13b], Hardisty 2022[59]) and (ii) the further clarification and specification of the FAIR principles with regard to the implementation of FAIR's key concept "machine-actionability" (Lannom 2021[60]) in the context of data management plans (Miksa 2019[8c]).

Taking up the concept of machine-actionable data management plans (maDMP; *Miksa 2019*), we envision DiSSCo's DMP not as a kind of static document but as a dynamic structure involving FAIR digital objects, which integrate all relevant information in the data life cycle.

Timeline and Summary Recommendations

Table 4. Data Management Plan Implementation (ICEDIG) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
GAP	010	"Machine-actionability", the capability of machines to handle data autonomously and appropriately is a core objective of FAIR and thus FDOs (Jacobsen 2020[7c]). The concept was i.a. developed to describe an aspired transformation of DMPs from essentially static documents towards an application-based tool chain integrating and automatising all relevant information steps in the data life cycle: the machine-actionable data management plan (maDMP; Miksa 2019[8d]). **Due to several constraints in DPP, a merge of the maDMP concepts and the FDO approach wasn't elaborated to the same extent as a prototype in DPP like the machine-learning and human-(crowd-) based annotation pipelines** (cp. Leeflang 2022[9c]; Grieb 2021[10d]). The objective and potential of the maDMP are the seamless integration of																	

		the reporting tools into the existing network of project tools and services to release involved humans (developers, scientists, curators etc) from the "annoying administrative exercise" (Miksa 2019) to write a DMP as a matter of form. It should be feasible to realise this level of automatisation for DiSSCo's Digital Specimen services, since the fundamental Digital Object Architecture provides already a network of FAIR-enabling services, which (prototypically) implement the policies, rules, procedures and infrastructures for specimen data during its whole life cycle. **The main task now is to provide a proper alignment on several levels (data model, services) with common specifications (DCAT, PROV-O) to make the DES more "speech-enabled" with regard to retrospective and prospective provenance.**														
Tr	004	**Further develop and refine the provenance model in openDS.**	I	I	I	I										
Tr	005	**Provide a more complete capture of digitisation processes in MIDS** by establishing a DES as output of a MIDS level 3 description (align MIDS information elements with openDS) with the aim to facilitate the transfer of digitisation workflows.	I	I	I	I										
H	006	**Establish a FDO type for a machine-actionable DMP** Link all relevant outputs of a DES' lifecycle to a maDMP-FDO comprising links to retrospective and where applicable prospective provenance objects, to related services like re3data for repository information, to machine-actionable policies in a structured format like ODRL. Elements of this FDO type must support a sufficient level of granularity (no long free text as bitstream) to enable reaggregation of information elements to adjust the DMP dynamically.	P	P	P	P	I	I	I	I	F	F	F	F			

			P	P	P	P	I	I	I	I	F	F	F	F					
H	007	**Implement linking to comparable service networks like ENVRI-FAIR** Enable the interoperability of DiSSCo's core entities involved in RDM by providing mappings to a common vocabulary shared among relevant data providers like the ENVRI-FAIR network. A common standard here is W3C DCAT, which is centered on requirements of data exchange of data portals in Europe providing semantic interoperability by reuse of (e.g. EuroVoc) and mappings to existing vocabularies (e.g. INSPIRE, schema.org[15c]). Predefined profiles are available like DCAT-AP used for EU infrastructures, EPOS DCAT-AP employed in ENVRI-FAIR which provides starting points for DiSSCO's mapping (or subclassing) of core entities like DES, collection, CMS, enrichment service to dcat:Resource, dcat:catalog dcat:Service etc.	P	P	P	P	I	I	I	I	F	F	F	F					
M	008	**Enable machine-actionable data management for DiSSCo by providing a FAIR ecosystem of DMP-related services** Data cannot be FAIR when an infrastructure does not implement policies, rules and procedures for FAIR. Accordingly, key services to support maDMPs have to be established involving automatic recommender services for licensing, tools to compose access and update policies using machine-actionable policy elements based on machine-actionable formats like PROV-O and ODRL, and services to enable the automatic forwarding of RDM tasks to curators and other agents when specific input is expected. This involves an assessment and check for suitability of available services like the EUDAT license tool[41c].	P	P	P	P	I	I	I	I	F	F	F	F					

M	009	**Implement prospective provenance for DES**	P	P	P	P	I	I	I	I	F	F	F	F				
		Provide a provenance model to capture the sequence of operations leading to a certain state of a DES and related biodiversity FDOs in the form of a "recipe" to facilitate restartability and partial updates from a selectable checkpoint. This should also enable the transfer of operation sequences and workflow parameters to data objects with similar constraints, e.g. the sharing of digitisation pipelines between similar subtypes of the DES. In DPP, a first prototype was developed to support "restartable DES" making use of JSON Patch[42c] and PROV-O which should be elaborated to support full workflow software environment preservation involving tools like CWLProv (Khan 2019[43c]).																

Discussion and Conclusions

Based on the results from preceding projects like ICEDIG, joint discussions in the openDS (Digital Specimens) breakout group and community events like the DiSSCo Prepare Round Tables or the Consultation on the Convergence of Digital and Extended Specimens[61], the following essential characteristics for the data management of Digital Extended Specimen were prioritised (cp. Leeflang 2022[17c]):

- Digital Specimen is the core component and the primary digital object type of the DiSSCo architecture

- Accuracy and authenticity of the Digital Specimen

- FAIRness (BioDT WP5[62a])

- Protection of data (legal regulations and community norms)

- Preserving readability and retrievability (BICKL WP7[63])

- Traceability (provenance) of specimens   (BioDT WP5[62b])

- Annotation history

- Determinability (status and trends) of digitisation (MIDS/SYNTHESY+ NA4.1[64])

- Securability (authentication, authorisation, accounting, auditing

The theoretical core of the Digital Extended Specimen concept – DES as a specific type of FDO (FAIR Digital Object) that acts as a digital surrogate for a specific physical specimen in a natural science collection (Islam 2022[65]) - is closely aligned with the objectives of the EU's Destination Earth (DestinE) initiative which is to develop an accurate digital model of the Earth via sets of thematic Digital Twins. Within the *Biodiversity Digital Twin for Advanced Modelling, Simulation and Prediction Capabilities* (ID 101057437) (BioDT WP5) project, DiSSCo partners are developing a FAIR ecosystem involving services for reproducible workflows, FDO type registration and mapping of semantic artefacts building on infrastructures of DiSSCo, ELIXIR, LifeWatch and others. In particular BioDT T5.4 *Supporting FAIR workflows and models* works on essential data management topics closely linked to DES like data plus method/workflow preservation, aggregation and sharing[66] based on the integration of CWLProv (Common Workflow Language) and digital objects.

DiSSCo partners (lead: Naturalis) will create a persistent identifier registration (PID), discovery, resolution and indexing service or DES with the Biodiversity Community Integrated Knowledge Library[67] that will enable data related to the specimens and samples to be unambiguously retrieved, re-used, linked, published and cited. The PID Infrastructure is of course an essential infrastructure for data management (in particular for compound/interlinked objects) in DiSSCo (cp Hardisty 2021[36b]).

With regard to processuality of digitisation, i.e. the actual process of making physical objects digitally available as Digital Extended Specimens, there are close interrelations of DiSSCo's data management and the Minimum Information about a Digital Specimen (MIDS[68]). MIDS development was part of SYNTHESYS+ and is continued as Biodiversity Information Standards (TDWG) MIDS Task Group. MIDS level 3 is intended to comprise all lower MIDS levels (plus additional information elements) about a digitised collection object encapsulated in the form of a machine-actionable DES (subclass like BotanySpecimen etc).

Method

· Within the framework of open Digital Specimen specification (openDS), specific classes and properties for the comprehensive provision of provenance data were designed and developed (ods:EventProvenanceRecord).

· Based on these classes, several prototypical use cases involving different agents and roles like machine-based annotation, term/taxonomic revision by curator, enrichment by scientist were implemented, and (preliminary) results were iteratively passed back in the modelling process.

- Subsequently, serialisations to capture retrospective provenance data including the details of executed processes to derive a specific result were developed involving standards like W3C PROV-O, the Web Annotation Ontology and (technically) JSON-LD.

- Furthermore, an experimental serialisation path employing JSON patch was drafted to provide prospective provenance information in form of the computational steps necessary to create particular results like an annotation. In terms of reproducibility this enables a digital object to be restartable.

3.3.3 Digital Specimen Data Model

DiSSCo Modelling Framework and Data Model

Description   *DPP D5.2*

DPP D5.3

The interoperability of DiSSCo RI software components and services highly depends on the availability of an agreed data model covering all aspects of specimen-related information, for example, specimen and observational data, taxon names, DNA sequences, and publications.

The existing landscape, of bio- and geodiversity data formats and standards, is heterogeneous and modelling approaches are hindered by the current use of different modelling workflows for similar or identical data domains, incomplete version histories, insufficiently documented concepts, and the lack of machine-readable links to related data definitions. The DiSSCo Modelling Framework will provide modelling capabilities for required groundworks in order to shape DiSSCo Digital Specimen Object Specifications (DSOS).

Specific use cases range from semantic annotation of collection-related research data to general semantically annotated DiSSCo APIs (application programming interface). Based on an existing technical platform (e.g. WikiData) and existing standards and data formats developed by the community (e.g. DwC-Darwin Core, ABCD-Access to Biological Collection Data, CDM-Clean Development Mechanism), the modelling framework will provide a mechanism for agreeing and documenting a core set of "DiSSCo data elements". The Platform will provide persistent ID for each concept, relations to elements in existing standards, as well as an API for machine-readable access to the data model.

The main objectives of this task are (1) a compilation of target resource (information) types, e.g. taxon names, publications, DNA sequences, (2) an analysis of existing data standards (e.g. DwC, ABCD, GGBN- Global Genome Biodiversity Network, Audubon core) according

to their matching capability with the required resource types, (3) identification of data standards forming the basis for the DSOS, (4) definition of specifications and technical implementation of the required modelling capabilities, e.g. WikiBase, and (5) an alignment of identified standards for building the DiSSCo-specific applications for digital objects.

Timeline and Summary Recommendations

Table 5. Digital Specimen Data Model Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	011	We foresee that the basic modelling processes will have been established with the end of the DiSSCo preparation phase. However, there will be the need for further coordination of works on the model itself. The DiSSCo synchronisation group for data standards strongly recommended allocating resources for the coordination of modelling activities. **Allocate resources for the coordination of further modelling activities and recruit a modelling expert.**	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F	F
M	012	**Coordinate a domain expert group for peer review of modelling decisions.**	P	P	I	I	I	I	I	I	I	I	F	F	F	F	F	F

Discussion and Conclusions

The modelling platform has been successfully deployed (see Deliverable 5.2, Digital Specimen Data Model[69]) and integrated into the DiSSCo technical architecture (see Deliverable 6.2, Implementation of DS Architecture[70]). However, developing the model content (i.e. concrete data types processed by DiSSCo) is an ongoing process. The process needs to be synchronised with related standardisation activities, for example in TDWG and SYNTHESYS.

Method

We have implemented a functioning first version of a Wikibase based platform for creating a DiSSCo data model with a focus on essential data types. In cooperation with Task 6.2

<u>Implementation of DS Arch</u> (Digital Specimen Architecture), there are regular meetings for defining the content of the DiSSCo Digital Object Specification (D5.3).

3.3.4 Semantic Annotation

Semantic enhancement and interoperability

Description   *DPP D5.4*

Semantic enhancement of existing textual collection information by linking to (external) authority files or other semantic resources is a key effort to improve discoverability, comparability, data quality and linkage of the digital objects. Furthermore, enabling semantic interoperability of collection data will facilitate a new generation of applications processing and connecting this information to the rapidly growing globally linked open data graph.

The main objective of this task is to provide construction plans for required basic components and document best practices for semantic enrichment at the collection level, starting with priority concepts such as taxa, persons (collectors) and geographic features.

Building on preliminary works of the CETAF ISTC (Consortium of European Taxonomic Facilities - Information Science and Technology Committee) as well as the OpenUp! project, this task will instigate a broad campaign for semantic enrichment activities in all DiSSCo member organisations and ensure that related activities are aligned with data management processes in CETAF member organisations.

Timeline and Summary Recommendations

Table 6. Semantic Annotation Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing														
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	013	**Develop tools that enable automated enrichment** both as part of the digitisation process and as part of publication pipelines.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
H	014	**Organise a working group to exchange experiences on enrichment, and discuss and promote curation of open resources** such as Wikidata to avoid double-work.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

M	015	During standards development, **push to support enrichment with multiple PIDs from different resources for a single concept.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
M	016	**Organise and promote manual enrichment activities at individual collections** as a part of capacity enhancement.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
M	017	During standards development, **push to support verbatim or raw data, so no loss of information occurs** as a result of enrichment.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

Discussion and Conclusions

In this deliverable, the basic components of a semantic enhancement process are explored through various questions, such as the current state of the data to be enhanced, the context in which the enrichment takes place and the desired outcome. The many different answers to these questions make it difficult to provide a generic construction plan. Instead, best practice recommendations are detailed and for common critical components the current landscape is explored. These components include

- PIDs and their resolution mechanism

- Enrichment protocols

- Data standards that need to support enrichment activities

PIDs are integral for enrichment to function. Discussions held during various workshops in the DiSSCo-linked COST MOBILISE Action underscored the need for increased adoption of PIDs at the source (preferably as close as possible to initial digitisation), and the urgent need to address some concepts for which PID systems are still not optimally functional or otherwise lacking, such as specimens, specimen citations and specimen media. This was re-emphasised more recently (e.g. in this report[71]) in the BiCIKL (Biodiversity Community Integrated Knowledge Library) project, where specimens are one of the many entities to be linked to other nodes of the Biodiversity Knowledge Graph.

In particular for the data types focused on in this report (people and geography), Wikidata shows a lot of promise for enabling FAIR semantic enrichment, both as a disambiguating PID broker and as a fallback resource for enrichment cases that are not well covered by existing authority resources. More dedicated resources exist, such as various authority files for people identities and public gazetteers of geographic features. These are not always openly accessible and updateable. Stability in the digital world is also not a given and is difficult to predict in advance. Mechanisms that enable PIDs to be minted and keep

resolving may end up suffering from a lack of maintenance, deprecated underlying technologies and/or run out of funding. Similar to other long-term endeavours with implicit uncertainty (e.g. forestry), portfolio approaches are to be recommended, where more than one PID can be utilised or at least resorted to in case the primary one fails.

Various **protocols**, algorithms and other tools are described in this report that already leverage Wikidata, as well as a few other resources such as Geonames and ORCID. A task in the BiCIKL project, 7.3[72], aims to build (semi)-automated workflows that facilitate enrichment of collection data on people and geography with Wikidata identifiers, that can be implemented by multiple parties and in particular the DiSSCo infrastructure.

Such workflows are also being prototyped in the SYNTHESYS+ JRA3 work package with an explicit focus on Machine Learning (see here[73] for the landscape analysis). While these algorithms can create considerable added value, they are not the ultimate solution and slower, human-in-the-loop approaches will continue to be necessary. Automated methods also impose stronger validation requirements and may lead to reconciliation or synchronisation conflicts, as the volume of enrichment annotations they generate is much greater.

Data standards are being developed to better support enriched data, but gaps still exist and uptake needs to be promoted. Development of the openDS standard in DiSSCo Prepare WP6 should accommodate these problems.

Method

The current status of semantic enrichment for people and geographic features was assessed, including pilot efforts for automated linking and ongoing developments regarding data standards. Based on these results, a breakdown was made for the different steps in enrichment workflows, taking diverse contexts into account when making recommendations for each step.

3.3.5 Data Mobilisation (Institutional)

Collate, refine and implement best practices for data mobilisation at the institutional level to develop the DiSSCo plan for data mobilisation and curation pipelines

Description   *DPP D3.2*

How do you best prepare collections for digitisation, digitise them, curate the associated data, publish this information and measure the outputs? What are the options and rationale for different types and sizes of collections, when should this be outsourced and what

different project management approaches are most appropriate in this range of circumstances?

This task seeks to address these questions, describing and refining best practices and building on a substantial investment from prior and current projects (MOBILISE COST Action, ICEDIG; SYNTHESYS+ - Table 4) and feeding these into DiSSCo Prepare <u>WP8 Stakeholder Engagement and Communication Strategy</u>). Consolidating what is known into a community-edited manual (supported by WP5), and other relevant platforms, WP3 streamlines the reuse and implementation of these procedures and enhances digitisation capacity across the DiSSCo collection-holding organisations.

Recommendations have been included in the final deliverable for the continued maintenance and development of the digitisation guides website during DiSSCo construct/DiSSCo operation phases. These include those in the following table.

Timeline and Summary Recommendations

Table 7. Data Mobilisation (Institutional) Recommendations

NB: ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	019	**Establish a Digitisation Interest Group** with representation and linkage to other relevant community groups (e.g., TDWG, CETAF DWG/ISTC) to continue development of the DiSSCo Digitisation Guides, and discuss challenges and solutions for all aspects of digitisation.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
H	022	**Determine how much, if any, investment DiSSCo plans to make in digitisation infrastructure and how much support will be provided to institutions to increase their digitisation capacity.** This includes support to keep the Digitisation Guides up-to-date and to foster community curation. This would ideally be combined with both in-person and online <u>training</u>, and potentially support planning and implementation of physical and digital infrastructure investment. We recommend focusing content and support on new or recently	P	P	P	P	I	I	F	F	O	O	O	O	O	O	O	O

		established national nodes where digitisation is a priority, and national nodes which are working on national or regional digitisation programmes.																	
M	018	**Establish a formal but light-touch process for updating the DiSSCo Digitisation Guides**[44c] **website**, reviewing and accepting new content (including workflows) to the site.	P	P	I	I	F	F	O	O	O	O	O	O	O	O	O	O
M	020	In collaboration with the Helpdesk, **provide a range of support channels to answer questions**, provide in-person and online training, and establish digitisation best practices across the DiSSCo community.	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O
M	021	**Assess user requirements of the guides website**, including UI improvements.	P	P	P	P	I	I	I	I	F	F	F	F	O	O	O	O

Discussion and Conclusions

There remains a gap in supporting digitisation activities in DiSSCo, especially for collections that are just starting to make their collections available. The DiSSCo Digitisation Guides[44d] website is a starting point for some collections, and follow-up work by MOBILISE[74] (Jan 2023) has increased guidance and content. DiSSCo will need to provide additional support to keep the Digitisation Guides up-to-date and to foster community curation. This would ideally be combined with both in-person and online training, and potentially support planning and implementation of physical and digital infrastructure investment. We recommend focusing content and support on new or recently established national nodes where digitisation is a priority, and national nodes which are working on national or regional digitisation programmes.

Current content on the DiSSCo Digitisation Guides website includes:

·   Nine Standard Operating Procedures (SOPs)/workflows, including herbarium, pinned insect and microscope slides (MS3.5)

·   Extract, Transform and Load (ETL) Best Practices (MS3.6)

·   Manual Transcription and Specimen Image Capture Guidance

·   Pre-Digitisation Curation Checklist (in development. MS3.7)

Although the final deliverable[75] for this task was completed in July 2022, the site was continuously maintained thereafter by NHM staff until the end of DiSSCo Prepare. NHM

staff intend on supporting the site after the formal conclusion of DiSSCo Prepare, but to be truly sustainable we need to establish a community of maintainers and contributors.

Method

We have built a community digitisation guide website (dissco.github.io[76]), which contains digitisation SOPs, ETL best practices, and a pre-digitisation curation checklist. A standard template was developed for SOPs, which includes the use of business process model and notation (BPMN) for workflow diagrams. GitHub was chosen as it can function as a light-weight content management system for publishing websites and managing permissions.

3.3.6 Taxonomic Backbone

Modernising technical infrastructure for science data mobilisation and publication

Description   *DPP D5.5*

The modernisation of key services, especially of services for data currently underdeveloped in the DiSSCo community, is of great importance for the overall improvement of the technical readiness level from the DiSSCo RI.

This aspect of the task focused on construction plans for the improvement of technical infrastructure in the identified key areas of taxonomic services. Construction plans for harmonising life science taxonomic checklist services and their integration into DiSSCo architecture need to be developed to exploit their full value as a taxonomic backbone. This task focused on the investigation of existing tools, specifically the existing Catalogue of Life[77](COL) services and the new Checklist Bank[78] suite of tools developed in conjunction with GBIF (Global Biodiversity Information Facility), and providing necessary plans for integration into the overall DiSSCo technical architecture.

Timeline and Summary Recommendations

Table 8. Taxonomic Backbone Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

H	023	**Focus on articulating more detailed requirements for taxonomic names services** *Species 2000 would recommend a further in-depth exploration with DiSSCo into the requirements for taxonomic names services. *Engage with COL to further explore the benefits that COL and ChecklistBank might leverage from interoperability with DiSSCo, and resource implications of supporting those requirements in the DiSSCo architecture.		P	I	F						
H	024	**Assess and refine the potential technical approaches to interoperability with ChecklistBank** *Use technical pilots to assess the possibilities and limitations of using COL's public services and open APIs to meet DiSSCo integration requirements. *Engage with COL to explore the potential implementation of the event-based interoperability model, and any associated investment that might be required into COL development.		P	I	I	I	F				
M	025	**Engage further with GBIF and COL in the context of the Alliance for Biodiversity Knowledge** *Continue to support the international effort by GBIF and COL to build one common infrastructure for taxonomic names through ChecklistBank. *Engage with GBIF and COL to support the implementation of PIDs for taxonomic names and concepts.	P	P	P	P	I	I	I	I	F	
L	026	**Engage with the TDWG Paleo Data Working Group on representing fossil taxonomy in the <u>Digital Specimen Data Model</u> and consult with COL on improvements in ChecklistBank.**							P	I	I	F

Discussion and Conclusions

DiSSCo's favoured technical model for sustainable interoperability with external services is based on the event publishing approach, originally proposed for integration with institutional collections management systems. An important message that emerged is that this would need investment and collaboration to achieve with Catalogue of Life. Catalogue of Life would expect DiSSCo to partake in the joint investment of the sustainability and maintenance of its services, including Checklist Bank, that will be jointly governed by international biodiversity data initiatives and infrastructures.

Accordingly, two high-level recommendations for the integration of geo-collections and geo-classification services arose from this task:

1. Apply a phased and pragmatic technical approach in working towards interoperability with external services.

   An event-based approach represents an efficient and scalable method for interoperability in the long term, but is also dependent on third parties prioritising the required development and having resources available, which may not be easy to achieve without investment of funds and/or technical resources from DiSSCo for that development.

   In the shorter term, a more effective approach may be for the DiSSCo development team to continue to pilot more bespoke integrations with Catalogue of Life. These quicker proofs of concept might also help to demonstrate the benefits of interoperability and strengthen the case for funding from external sources to work towards the more robust event-based architecture.

2. Engage more deeply with Catalogue of Life and the Checklist Bank programme through the Alliance for Biodiversity Knowledge.

   Although DiSSCo may make use of COL's open tools for the community, there is greater long-term potential in developing a collaborative relationship with COL to better understand and support the needs for taxonomic services across the DiSSCo user base. There is also an expectation from Species 2000 that major users, like DiSSCo, that have dependencies on COL Checklist and Checklist Bank services will help to carry the financial sustainability for Checklist Bank and jointly maintain it as a global resource. COL will invite DiSSCo, just like GBIF and other biodiversity data infrastructures, to become part of the governance of the Checklist Bank infrastructure. Besides the strategic considerations, investment from DiSSCo is likely to be required if there is any expectation of development work in COL to pave the way towards the more robust event-based model of interoperability with Checklist Bank.

Method

Reviews of the Catalogue of Life and ChecklistBank were carried out through a combination of desk-based research, document review and direct input from developers and representatives of the platforms. User requirements were extracted from earlier DiSSCo outputs, in particular the user stories analysed in Task 1.1, and further detail was added to the picture by running an event-storming workshop, which engaged wider members within the community to help to identify and prioritise those events in external systems that would need to be reflected in the data within the DiSSCo architecture. Technical requirements from architectural design and pilot activities carried out by the DiSSCo developer team were also incorporated, with a focus on an exploration of the Catalogue of Life API capabilities and incorporating the taxonomic lookup services into automated workflows within the DiSSCo enrichment services. From these activities, it was possible to make a gap analysis between the technical capabilities and resource capacity of the reviewed platforms on one hand, and DiSSCo's requirements and technical roadmap on the other, which in turn formed the basis for a set of recommendations for DiSSCo Construct.

3.3.7 Knowledgebase

DiSSCo Knowledgebase for technical development

Description   *DPP D5.1*

DiSSCo Prepare builds on profound technical knowledge from various sources and initiatives. In order to allow for efficient knowledge and technology transfer for partners building the DiSSCo technical backbone, a central and freely-accessible DiSSCo KB (Knowledgebase) has been designed and implemented.

The DiSSCo KB will provide (1) structured technical documentation of identified DiSSCo technical building blocks, such as web services, PID systems, controlled vocabularies, ontologies and data standards for bio- and geo-collection objects, collection descriptions, digital assets standards as well as domain-specific software products for quality assurance and monitoring, (2) an assessment of their technical readiness for DiSSCo, as well as (3) specifications on their relevance for the overall DiSSCo technical infrastructure and the DiSSCo data model.

This will be achieved together with the international community (data) standards organisations (TDWG, GGBN).

Timeline and Summary Recommendations

Table 9. Knowledgebase Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	027	**Establish a dedicated person (editor) or editorial team** responsible for quality control, content management and curation of the DiSSCo KB.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
H	028	**Engage or provide technical maintenance services and basic software support** for KB.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
H	029	**Further, integrate with DiSSCo's Authentication and Authorization Infrastructure** (AAI).			P	P	P	I	I	I	I	I	F	F	O	O	O	O
H	030	**Further, implement permanent ongoing curation, maintenance and development** (based on future requirements) of the service to remain a trusted information source.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
M	031	**Develop training and training materials** for users of the DiSSCo KB.			P	P	P	I	I	I	I	I	F	F	O	O	O	O
M	032	**Further, closely link or integrate DiSSCo Helpdesk** (e. g. JitBit)			P	P	P	I	I	I	I	I	F	F	O	O	O	O
L	033	**Further, integrate DiSSCo Training software** (e. g. Moodle)			P	P	P	I	I	I	I	I	F	F	O	O	O	O
L	034	**Further, research and integrate further functionality upon request.**			P	P	P	I	I	I	I	I	F	F	O	O	O	O

Discussion and Conclusions

The DiSSCo Knowledgebase aims to serve as a central hub for technical documentation, research output and further resources relevant for the DiSSCo and wider community. As such, close links and dependencies exist with the DiSSCo Helpdesk and other DiSSCo eServices such as the Authentication and Authorization Infrastructure (DiSSCo AAI layer).

Method

DiSSCo Prepare WP5, Tasks 5.1, in close collaboration and exchange with other DPP project partners, collected the extent of information types expected to be stored in the Knowledgebase. To get a more complete picture, this was also discussed together with project overarching bodies such as the DiSSCo CSO and Technical Team. A comprehensive landscape analysis was made to select a system as document repository, based on the list

of top priority criteria for the decision of an appropriate software for the Knowledgebase. According to the requirements and with respect to the implementation and maintenance as well as latest releases, the most promising system was DSpace. A first beta version was made available and introduced during the First virtual All Hands Meeting (AHM) of DiSSCo Prepare (January 18-22, 2021) for presenting the functionalities to the audience, in order to discuss possibilities to structure the content, and to collect constructive feedback and requirements from DiSSCo partners. After adjusting the Knowledgebase with the feedback received in the AHM, a virtual meeting (DiSSCo Prepare Round table "Organization of knowledge and documentation for stakeholders") took place on July 6th, 2021 with more than 60 participants from the DiSSCo Prepare project and other interested parties. It was held to identify the stakeholder groups that need access to knowledge and documentation associated with the DiSSCo RI, to discuss how to organise the knowledge in such a way that optimal access is provided for these stakeholders, and to identify resources (existing elsewhere or missing) that should be provided. The pilot implementation with the DiSSCo branding design, current content and functionalities are available at the domain: https://know.dissco.eu/[79].

3.3.8 Helpdesk

Helpdesk and user support services

Description   *DPP D2.2*

DiSSCo needs to be in full and constant alignment with the needs of its identified users. This task defines the DiSSCo Helpdesk, that will provide the necessary information on the use of the infrastructure, the services and facilities offered. The DiSSCo Helpdesk will be a vital tool for increasing satisfaction of external users and stakeholders with the infrastructure, ultimately strengthening DiSSCo sustainability.

Timeline and Summary Recommendations

Table 10: Helpdesk Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing														
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

H	035	**Offer three complementary services for DiSSCo Helpdesk: FAQs - Frequently Asked Questions; Human Intermediated Question and Answer Service; and Supporting Documentation.** With respect to supporting documentation, most DiSSCo services are still under development and, therefore, production of supporting documentation should be considered an ongoing process.	P	P	P	P	I	I	I	I	F	F							
H	036	**For DiSSCo Helpdesk use the JitBit Helpdesk Ticketing System adopted for ELViS helpdesk.** Considering that the DiSSCo community is already familiar with this system, we anticipate no major issues, and no further preparation activities are needed to initiate its construction in 2024. We are confident that its construction will be finished by 2025.	P	P	P	P	I	I	I	I	F	F							
H	037	**Integrate DiSSCo Knowledgebase (KB) with DiSSCo Helpdesk** which will provide links to new and/or relevant content; automatically generate search results based on incoming questions, supporting personalised responses; and linking to other DiSSCo eServices, etc.	P	P	P	P	I	I	I	I	F	F							
H	038	**Follow the recommendations of the Web Accessibility Initiative (WAI) for the Helpdesk**, embracing an inclusive design that meets the needs of those who are permanently or temporarily disabled.	P	P	P	P	I	I	I	I	F	F							
H	039	**Allow real-time monitoring of the support process** in The Helpdesk, namely the Human Intermediated Q&A Service. **Develop a Service Level Agreement** (SLA), in which goals are set to meet users' expectations.	P	P	P	P	I	I	I	I	F	F							
H	040	**Provide training to support staff on the policies and procedures, as well as on the tools and technology** they	P	P	P	P	I	I	I	I	F	F							

		will be using. This will help ensure that the Helpdesk team is able to provide high-quality support to users.															
M	041	**The Helpdesk should be multilingual,** offering the researcher the possibility to interact with the system using native language statements/commands. Multilingualism should be fully implemented in FAQs and in support documentation, and although English will be the main working language of the Human Intermediated Q&A Service, it should have the possibility to receive questions and provide answers in different languages. Automatic detection and translation of languages should be considered.	P	P	P	P	I	I	I	I	F	F					
M	042	**Develop supporting documentation.** Most DiSSCo services are still under development and, therefore, production of supporting documentation should be considered an ongoing process.	O	O	O	O	O	O	O	O	O	O	O	O	O		

The Helpdesk will use the JitBit Helpdesk Ticketing System adopted for ELViS[80], so we anticipate no major issues, and no further preparation activities are needed to initiate its construction in 2024. We are confident that its construction will be finished by 2025.

Discussion and Conclusions

Three different modalities of interaction with the user are envisaged: **FAQs** - Frequently Asked Questions; **Human Intermediated Question and Answer Service**, which will use a support ticket system and will include three lines of support; and how-to guides and other **support documentation**, that will offer comprehensive information and assistance, empowering the DiSSCo community to use services independently.

We expect that the parties who express interest in becoming future DiSSCo service providers will provide specialist support agents, who will constitute the second line of support of the Helpdesk Human Intermediated Q&A Service, for the specific service. Most DiSSCo services are still under development and, therefore, the constitution of the support teams is dependent on the development of respective services. The same is true for the production of FAQ and supporting documentation, which are also dependent on services development.

Prioritisation to make documentation available should take into consideration several factors like the size of the user group, the amount of usage, and the user-friendliness of the service. The questions issued through the Human Intermediated Q&A Service will be an important tool to inform about the demand concerning supporting documentation.

Multilingualism of Helpdesk is expected by the DiSSCo users, and, ideally, the Helpdesk should offer the possibility to interact with the system using native language statements/commands. This possibility will promote equal opportunities among users, and attract non-English speaking users and stakeholders, ultimately increasing satisfaction with DiSSCo and sense of community.

The Helpdesk should follow the recommendations of the WAI (Web Accessibility Initiative), minding everyone who is permanently or temporarily disabled.

Finally, the Helpdesk should be flexible and customisable, allowing the change of its structure in time so that it best fits users' needs, and accommodates new services.

Training of support staff will be necessary to ensure that the Helpdesk team is able to provide high-quality support to users. Such a need has been identified in Task 2.1, concerning DiSSCo Training Strategy.

Method

As a first approach for the identification of support services to be provided by DiSSCo Helpdesk, we looked into helpdesk services provided by other research infrastructures, namely:

- CORBEL Shared Services for Life-Science[81]

- DARIAH-EU[82]

- CLARIN[83]

- LifeWatch[84]

We have delineated the Human intermediated Question and Answer Service built on the top of work developed under the SYNTHESYS+ project, which has focused on supporting the use of ELViS, that will consequently manage Transnational Access (TA) and Virtual Access (VA).

In order to gather information on the needs of supporting documentation, we have developed two independent questionnaires, one directed to those responsible for the DiSSCo services and the other directed to users. These questionnaires were conducted via an online, by a customised Google form. The issue of multilingualism was also addressed in survey 2, as interviewees were asked about the importance of having Helpdesk services in their national language (if different from English).

04 System Architecture

4.1 Overview
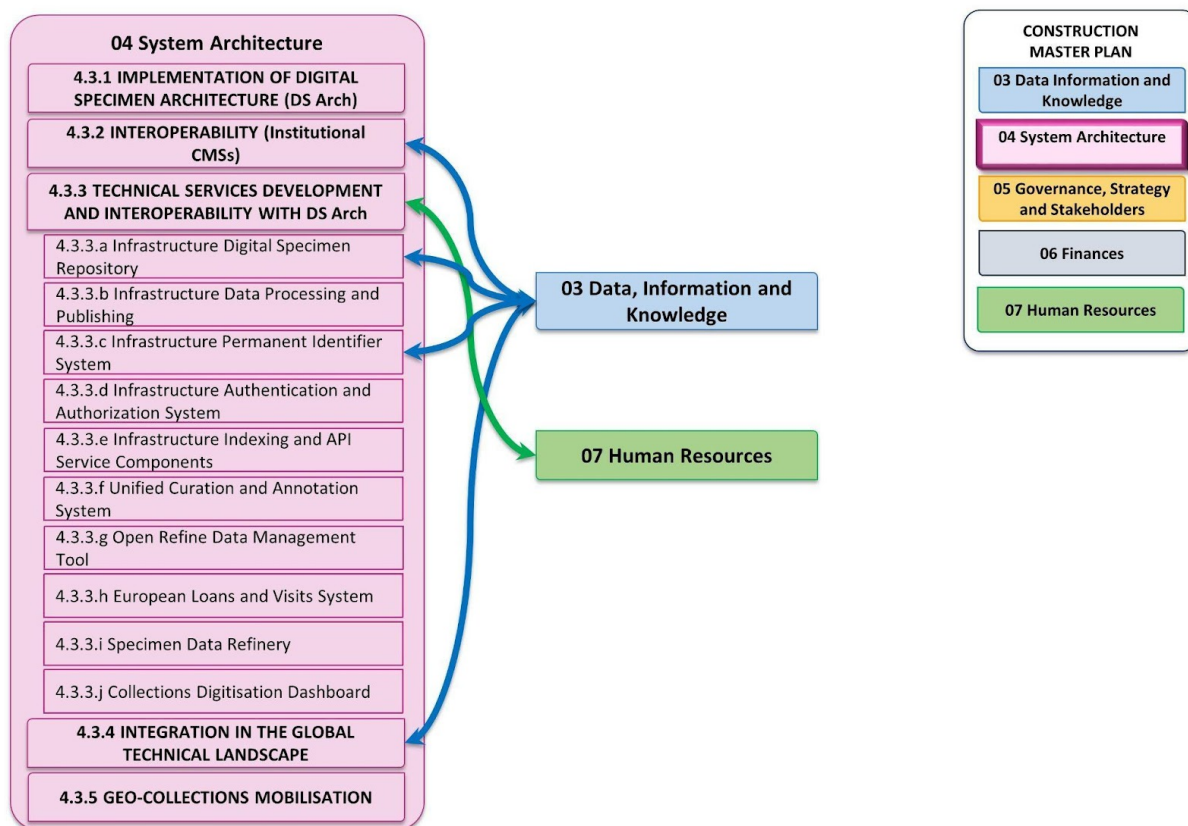
Figure 3. Synoptic graph showing the dependencies (arrows) between tasks in this chapter with other tasks or chapters (systems) in DiSSCo Construct.

DiSSCo's System Architecture based on DSArch, the Digital Specimen Architecture, is the key technical component that will support the data, information, and knowledge component of the DiSSCo RI. It needs to support mass digitisation and to provide a FAIR (Findable, Accessible, Interoperable and Reusable) digital representation of specimen data based on DiSSCo's Digital Specimen concept through integration with local Collection

Management Systems (CMS), and data processing and annotation services. It should be capable of storing and delivering data and providing technical services for DiSSCo's end-user services such as ELViS (European Loans and Visits System), UCAS (Unified Curation and Annotation System) and CDD (Collection Digitisation Dashboard). Following the Architecture recommendations provided by the ICEDIG project it needs also to be machine actionable to be 'Fully AI Ready' for future services using Artificial Intelligence (AI), for which the architecture design is based on the novel concept of FDO (FAIR Digital Object).

This chapter aims to provide:

- A detailed overview about the different activities towards the System Architecture that have been carried out during the preparatory phase of DiSSCo;

- The methodology and activities to further pilot and develop the System Architecture during the implementation phase;

- The methodology and activities to further develop and integrate end-user services with the System Architecture;

- Recommendations to guide the implementation phase towards the operation of DiSSCo.

- Recommendations to embed the System Architecture in the Global Technical Landscape and to promote the further development of geo-collection services.

The chapter describes the time path and further implementation steps of the DSArch which was prototyped in the preparatory phase to model the DiSSCo operational capabilities in line with the recommendations made in the ICEDIG project. It describes actions to continue prototyping by expanding and further implementing the open Digital Specimen (openDS) specification, which describes the Digital Specimen Data Model, ontology and API (Application Programming Interface).

Also included in the chapter is a description of the challenges and requirements towards interoperability and integration of DSArch with CMSs and the further implementation of technical services to support end-user services such as ELViS, UCAS, CDD and an experimental data provision tool for small collections based on OpenRefine. These technical services include the Authentication and Authorization Infrastructure (AAI), Processing, Indexing and Publishing services, Persistent Identifier (PID) Infrastructure, APIs and Specimen Data Refinery (SDR) services to support digitisation.

Other System Architecture aspects described are recommendations towards embedding DSArch in the Global Technical Landscape with a special focus on Geo-collections mobilisation and the actions that need to be taken to further develop and integrate the end-user services that make use of the system architecture and have been prototyped in SYNTHESYS+ and DiSSCo Prepare, to transform these into fully operational services to support the start of DiSSCo's Operational Phase.

4.2 Timeline and Summary Recommendations

Table 11. System Architecture Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

				"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing																
	Pri	Rec		2023				2024				2025				2026			
Task Name	**Typ**	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	Tr	043	(Before 2024): **Finalisation of a first release of the openDS specification** (see section "Discussion and Conclusions" for details). Immediate goal is a first release within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	Tr	044	(Before 2024): **Develop a coherent digitisation strategy binding for DiSSCo partners** including set up of infrastructures like sufficient storage capacities and the implementation of specifications (openDS, MIDS) to enable reliable feed into DiSSCo's core data pipelines.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

4.3.4 Integration in the Global Technical Landscape	Tr	093	**Finalise the first release of the openDS specification** (a technical necessity, see section "Discussion and Conclusions" in CMP_T6.4[11b] for details) within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.4 Integration in the Global Technical Landscape	Tr	094	**Ensure close involvement and alignment of DS/DS architecture with developments of Digital Object Architecture, FAIR Digital Objects Forum (FDOF) FDOs including FDO typing, machine actionability, and FDO profiles and attributes.** As noted above, the development of DiSSCo's Digital Specimen concept involving the corresponding service architecture is highly leveraging on Digital Object Architecture. DOA and in particular FDOs as core data models are currently the subject of intensive review and (re-)specification process. The procedure is documented in the FDOF Public Document Management Register[12b]). Of particular importance for the DS/DS architecture are the specifications of FDO typing (important with regard to subclassing DS, see subclasses proposed in Addink 2022 like BotanySpecimen, MycologySpecimen, GeneticResourcePlantSpecimen etc.), Machine-Actionability (the core objective of the FAIR principles to realise a global data ecosystem that can be navigated independently by machines), and FDO PID Profiles and Attributes (which provides essential Kernel Attributes of the FDO when an FDO's PID is	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

			resolved, cp. Islam 2020[13c]). It is fundamental for DiSSCo's technical architecture to be fully involved in this process and to ensure a close alignment with delivered results. Currently, members of DiSSCo's technical team are involved in the drafting process of all aforementioned documents with particular relevance for DiSSCo DS architecture.																	
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	045	**Transfer pilot plans of DiSSCo's service hub (e.g. repo[16b], data pipelines) detailed in D6.2 Implementation and construction plan to full production facilities** (Leeflang 2022[17d]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	046	**Develop, coordinate and operate a distributed, cross-institutional service support network to facilitate the local implementation of necessary technical adjustments linked to the Digital Specimen architecture.**	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	H	047	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation **[H] (concept: "UCAS - Unified Curation and Annotation System")** and annotation [M-L] (prototypes Specimen Data Refinery Hardisty 2022[18d], Machine Learning as a Service (MLaaS) pilot Grieb 2021[10e]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

4.3.2 Interoperability (Institutional CMSs)	H	051	**Install a local dedicated CMS person (developer) responsible for development to implement specific response procedures triggered by new event types.** For small to medium sized museums, we recommend hiring a developer at least for the first three months or for the period agreed upon between the institution and the professional to do the implementation requested by this event-driven API on the local CMS.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
4.3.2 Interoperability (Institutional CMSs)	H	052	**Install a local CMS technical team for ensuring compatibility** while both the CMSs and DiSSCo RI evolve.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
4.3.2 Interoperability (Institutional CMSs)	H	053	**Provide technical maintenance services and basic software support.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
4.3.2 Interoperability (Institutional CMSs)	H	054	**Further, develop, coordinate and operate a cross-institutional DiSSCo CMS support network (DiCSN).** A DiCSN is required to embed local CMS infrastructures of participating institutions in DiSSCo's Digital Specimen architecture. This will be based on the implementation of the API guidelines detailed in DPP D6.1 Harmonisation and migration plan for the integration of CMSs into the coherent DiSSCo Research Infrastructure (Glöecker et al. 2022[19b]) in close alignment with further development of the ingestion data pipelines outlined in DPP D6.2 Implementation and construction plan of the DiSSCo	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

core architecture (Leeflang 2022[17e]). As a major outcome of D6.1, detailed API guidelines for the bi-directional interfaces between local CMSs and the DiSSCo core infrastructure were elaborated. These guidelines could be either implemented as an extension of existing application interfaces to set up DiSSCo specific endpoints in CMSs that are web-based or by provision of a light-weight wrapper, a technique for extracting the data from documents that belong to non-web-based CMSs. To keep the technical integration of the local CMSs synchronised with the ongoing development of the retrieval infrastructure is of major importance to (i) enable continuous data feed into the core architecture (translator services, see D6.2) and to (ii) ensure the reverse data flow back from distributed curation into the local CMSs.

4.3.3 Technical Services Development and Interoperability with DS Arch	H	058	Find maintainers for the core e-services and plan steps to transform into future DiSSCo service providers.	P	P	P	I	I	I	I	I	I	F	F	F			
4.3.3 Technical Services Development and Interoperability with DS Arch	H	059	Develop the e-services into mature products for operation and integrate them with the DiSSCo FAIRDO infrastructure	P	P	P	I	I	I	I	I	I	F	F	F			

Section	H	No.	Description	1	2	3	4	5	6	7	8	9	10	11	12	13
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	H	062	**Implement Digital Specimen Repo as a component within the DiSSCo Arch.** Tracked in DD-132[20d].	P	P	P,I	P,I	P,I	F							
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	H	063	**Adopt the openDS data model.** Tracked in DD-133[21b].		P	P	P,I	P,I	F	F						
4.3.3.b Data Processing and Publishing Infrastructure	H	066	**Ensure data processing and publishing functionality** as part of the deployment of the Digital Specimen Repository (tracked in DD-132[20e])	P	P,I	P,I	P,I	F								
4.3.3.c Persistent Identifier (PID) Infrastructure	H	067	**Build a roadmap for DOI implementation** (from Handle to DOI) Tracked in DD-5[22c].		P	P	P	P,I	P,I	F						
4.3.3.c Persistent Identifier (PID) Infrastructure	H	068	**Create a service management plan for the PID service.** Tracked in DD-113[23b].	P	P	P,I	P,I	F								
4.3.3.d Authentication and Authorization Infrastructure (AAI)	H	070	**Finish User Profile development work and deploy it** in the current sandbox. Tracked in DD-13[24b].	P,I	F											
4.3.3.e Indexing and API Service Components Infrastructure	H	074	**Ensure API functionality** as part of the deployment of the Digital Specimen Repository[25c] (tracked in **DD-132**[20f]).	P	P	P,I	P,I	F								

Section		ID	Task	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
4.3.3.e Indexing and API Service Components Infrastructure	H	075	**Provide API documentation.** Tracked in DD-135[26b].			P	P,I	P,I	F								
4.3.3.f Unified Curation and Annotation System	H	076	**Finish UCAS Enhancement.** Tracked in DD-26[27c].	P,I	P,I	F											
4.3.3.f Unified Curation and Annotation System	H	077	**Consult users and redesign with UI/UX expert.** Tracked in DD-18[28b].		P	P	I	F									
4.3.3.h European Loans and Visits System (ELViS)	H	083	**Fix and improve current ELViS code base** (ELViS Code Evaluation[29b]). The Epic (DD-16[30d]) (as of Dec 2022) has several tickets in different states of completion.	I	F												
4.3.3.h European Loans and Visits System (ELViS)	H	084	**Align and integrate ELViS with the DiSSCo core infrastructure development** activities **and data management plan.** Also tracked in (DD-16[30e]).	P	I	I	F										
4.3.3.i Specimen Data Refinery (SDR)	H	087	**Align SDR with the DiSSCo core infrastructure development** activities. Tracked in DD-6[31b].	P,I	P,I	P,F											
4.3.3.j Collections Digitisation Dashboard (CDD)	H	090	**Automate import of digitised and collection holdings data into CDD** (note: this involves more effort from the collection holdings institutes than the DiSSCo technical team, and will therefore depend on the local technical team, their resources and training).	P	P	P,I	P,I	I,F	I,F	I,F	I,F	I,F	I,F	I,F			

				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
4.3.3.j Collections Digitisation Dashboard (CDD)	H	091	**Align CDD with TDWG standard** (Latimer Core[32c], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). This is related to DD-119[33c] and DD-32[34b].				P	P	P	I	I	I	F	F	F					
4.3.4 Integration in the Global Technical Landscape	H	095	**Develop the infrastructures for a FAIR service ecosystem leveraging FAIR Digital Objects for the greater range of Biodiversity, Life and Earth System Sciences. Provide the ecosystem of key services for FAIR Digital Objects** (Collins 2018[35b]) like FDO type and operation (for machine agents) registries, services for minting PIDs (cp. Hardisty 2021[36c]), repositories for the FDOs not only for the narrower target group of DiSSCo (NSCs) but also for the greater range of Biodiversity, Life and Earth System Sciences. The fundamental service ecosystem will be supplemented by community services like tools for modelling of new FDO types (for the diversification of DS subtypes see Addink 2022[37c]) or terminology services (TS) for the curation of semantic artifacts, including aforementioned mappings e.g. of DS concepts to classes of the new GBIF data model (a blueprint for a TS embedded in a FDO layer was drafted with participation of DiSSCo as Flexible Semantic Mapping Framework/SEMAF; Broeder 2021[38b]).	P	P	P	P	I	I	I	I	I	I	I	F	F	F?	F?	F?	F?

Task	Pri	#	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
[4.3.5 Geo-collections Mobilisation](#)	H	098	In collaboration with TalTech, MfN, the GeoCASe Advisory Board and the CETAF Earth Science Group, **continue to develop the joint vision for the development of GeoCASe,** explore models and opportunities for funding and resourcing GeoCASe development, and assess the overlap with GBIF in the palaeontology domain.	P	I	I	F												
[4.3.5 Geo-collections Mobilisation](#)	H	099	**Focus initially on publishing data to GeoCASe** through the preparation and presentation of Darwin Core Archives that GeoCASe can ingest.					P	I	F									
[4.3.5 Geo-collections Mobilisation](#)	H	100	**Incorporate support for Mindat classification and locality identifiers into the openDS specification.**					P	I	F									
[4.3.4 Integration in the Global Technical Landscape](#)	H-M	096	**Employ the developed FAIR ecosystem in the Bio*, Life and Earth System Sciences to enable advanced modelling, simulation and prediction capabilities.** This implies extending the focus from DiSSCo's core activity of NSC data management with the key feature of distributed curation during the lifecycle of the DS towards greater activities to support as well (derived) modelling lifecycles. Involved is the creation of new FDO types e.g. for machine-observations, -predictions and biologging plus an even deeper integration of tools and standards for portable workflows like the Common Workflow Language ([CWL](#)[39b]) and Research Object Crate ([RO-Crate](#)[40b]). This will build upon	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

			experiences from the Specimen Data Refinery prototyping in SYNTHESYS+ (Hardisty 2022[18e])																
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	M	048	**Situate DiSSCo as the major driver for the transformation of NSCs and other biodiversity data providers into integrated hubs for geo- and biodiversity data as part of European and global hyper infrastructures (EOSC, Gaia-X, GOSC) leveraging concepts like the FAIR Digital Twins** (Schultes 2022[45b]). This implies that the current technical focus of DiSSCo be widened by transferring and extending the established methods from collection-based research to other fields of geo- and biodiversity research (e.g. macroecology, biologging). Specifically, this means active provision and support of a FAIR ecosystem of services for the geo*- and bio*-communities involving infrastructures for modelling of new data types (beyond the current focus of the DS), hosting of registries for profiles and operations of those types and provision of service infrastructure for fundamental requirements like PIDs and AAI (cmp. section 5.3. in Wittenburg 2022[46b] for the high-level of FDO-compliance achieved by DiSSCo services).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.2 Interoperability (Institutional CMSs)	M	055	**Develop training and training materials** for the local CMS technical team.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

Section		No	Description																
[4.3.2 Interoperability (Institutional CMSs)](#)	M	056	**Further, foster the creation of an ecosystem of interoperable open-source CMS modules and services across the CMS community beyond DiSSCo.** Based on the structures and tools developed within the framework of DiSSCo, the wider community of open CMS developers will be addressed to adopt support for DiSCN-linked interface standards (see D6.1 above) and data models with special emphasis on Digital Extended Specimens and various DES subtypes proposed for openDS ([Addink 2022](#)[37d]) as well as for DES/Digitisation-linked standards with special emphasis on [MIDS](#)[47b] (MIDS level 3 is closely aligned with openDS/DES.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
[4.3.3 Technical Services Development and Interoperability with DS Arch](#)	M	060	**Establish service level agreements with identified service providers**	P	P	P	I	I	I	I	I	I	F	F	F				
[4.3.3 Technical Services Development and Interoperability with DS Arch](#)	M	061	**Establish FDO infrastructure compliance requirements for DiSSCo services**	P	P	P	I	I	I	I	I	I	F	F	F				
[4.3.3.a Digital Specimen Repository (DSR) Infrastructure](#)	M	064	**Align DSR with FDO specifications.**		P	P	P,I	P,I	P,I	F	F								

Requirement	M	No.	Description	1	2	3	4	5	6	7	8	9
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	M	065	**Integrate DSR with the DOI ecosystem** (see PID Infrastructure) Tracked in DD-5[22d] (this has dependency on a few global partners).	P	P	P,I	P,I					
4.3.3.c Persistent Identifier (PID) Infrastructure	M	069	**Ensure FDO compliance** by including Type Registry, FDO Record, and FDO Profile. Tracked in DD-10[48b] (this is also relevant for Digital Specimen Repo[25d]).		P	P	P	P,I	P,I	P,I	P,I	F
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	071	**Implement trust model**[49b] (related to UCAS enhancement work tracked in DD-26[27d]).	P	P	P,I	P,I	F	F			
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	072	**Seek guidance and help from EOSC and EOSC Life** expert resources via GRNet **to build a robust AAI infrastructure.**		P	P,I	P,I, F					
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	073	**Deploy and use AAI system** for all different DiSSCo services.				P	P	P	P,I	F	F
4.3.3.f Unified Curation and Annotation System	M	078	**Align with openDS data modelling work** for the annotation data model.		P	P	I	F				
4.3.3.f Unified Curation and Annotation System	M	079	**Align with FDO specification.**		P	P	P	I	I	F	F	

Reference		No	Description								
4.3.3.f Unified Curation and Annotation System	M	080	**Implement automation and machine annotation to speed up the annotation process** (needs some interaction with the SDR and other AI/ML development activities in the community).			P	P	I	I	F	
4.3.3.g Open Refine Data Management Tool	M	081	**Conduct surveys and consultation with users from organisations that do not have CMS**.			P	P,I	P,I	F		
4.3.3.g Open Refine Data Management Tool	M	082	**Provide a feedback plan to feed further development work.**					P	P	P,I	F
4.3.3.h European Loans and Visits System (ELViS)	M	085	**Align ELViS with TDWG standard** (Latimer Core[32d], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). Tracked via DD-119[33d].		P	I	I	F	F		
4.3.3.h European Loans and Visits System (ELViS)	M	086	**Ensure coordination and collaboration of distributed DiSSCo core architecture and ELViS code development (potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and ELViS core code base separately. Each of these teams will have different working cultures and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address incompatible workflows. It is also important	P	P	I	I	I	F		

			to identify different milestones aligned with the feature requirements and SRL. Working in such a distributed manner might require training and workshops. The use of agile product roadmap[50b] methodology is highly recommended.	P																		
4.3.3.i Specimen Data Refinery (SDR)	M	088	**Integrate SDR with RO-Crate, Galaxy and FDO specification** (alignment with the BioDT[51b] and BGE[52b] project as RO-Crate might be used there as well).			P	P	P,I	,F	F												
4.3.3.i Specimen Data Refinery (SDR)	M	089	**Provide training for local distributed technical teams in distributed ways of working (Potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and SDR core code base separately. Each of these teams will have different working culture and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address this problem. Working in such a distributed manner might require training and workshops.	P	P	P	I	I	F												
4.3.3.j Collections Digitisation Dashboard (CDD)	M	092	**Align CDD with the DiSSCo core infrastructure development** activities. Related to DiSSCo DD-16[30f].				P	P	P	I	I	I	F	F	F						

				C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
[4.3.4 Integration in the Global Technical Landscape](#)	M	097	**Assert DiSSCo's role as the major driver of the converging and harmonisation process of the Digital Extended Specimen with respect to establishing a joint digital object approach. Ensure compatibility of data models within approaches** aiming to provide digital object-based information entities like DiSSCo's Digital Specimen and BCoN's Extended Specimen with wider interoperability standards. It is significant that in particular outcomes of the Digital Extended Specimen conversion discussions ([Webster 2021](#)[53b]) comply with edge conditions set by overarching FDOt standards (e.g. on FDO Typing).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
[4.3.5 Geo-collections Mobilisation](#)	M	101	**Explore a potential relationship with Mindat for automated integration between DiSSCo and [mindat.org](#)[54b].** • Collaborate with Mindat to develop a specification and resourcing estimate for integration using DiSSCo's event-driven approach • Investigate potential use cases and appetite for the integration or linkage of DiSSCo Digital Specimen data in the Mindat platform		P	I	I	I											
[4.3.5 Geo-collections Mobilisation](#)	M	102	**Work closely with GeoCASe on an early proof of concept for event-driven interoperability with an external service** Use the process to:					P	I	I	F								

			• develop a blueprint and set of minimum requirements for external services to interoperate with DiSSCo using the event-driven architecture; • develop a detailed specification, resource requirements and costings for GeoCASe development to automate ingestion from DiSSCo using the event driven approach.																	
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	M-L	049	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation [H] (concept: "UCAS - Unified Curation and Annotation System") and **annotation [M-L]** (prototypes Specimen Data Refinery Hardisty 2022[18f], Machine Learning as a Service (MLaaS) pilot Grieb 2021).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)	L	050	**Provide a discovery service to cluster/match types of Digital Specimens (and other biodiversity Digital Objects) against requirements for scientific data.** Objective is to enable the determination of semantic similarity of DS and the suitability of the encapsulated data by providing (or ranking) of appropriate Digital Objects using DS attributes transformed to vector representations/embeddings (Le 2014[55b], Devlin 2019[56b], Wittenburg 2022 in sub.)	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

Section	L	#	Description																
4.3.2 Interoperability (Institutional CMSs)	L	057	**Further, evaluate, develop and integrate further functionality** upon request.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
4.3.5 Geo-collections Mobilisation	L	103	**Implement the changes specified in the GeoCASe early proof of concept** if appropriate resources can be made available.									P	I	I	F				

4.3 Task Summaries

4.3.1 Implementation of the Digital Specimen Architecture (DS Arch)

Evaluation of the DiSSCo Architecture

Description   *DPP D6.2*

Evaluating and testing the DiSSCo Architecture recommendations provided by the ICEDIG project to add further detail, resulting in an implementation and construction plan for the core DiSSCo technical infrastructure.

The infrastructure was prototyped to model DiSSCo operational capabilities and implement the core virtualisation infrastructure, supported by bi-directional interfaces from/to the DiSSCo facilities as specified in Task 6.1. The infrastructure will need to serve digital specimens and their related data for the DiSSCo scientific use cases as collected in ICEDIG. Required hardware infrastructure for storage and computing was described as well as integration with the European Open Science Cloud (EOSC). The task includes the following activities:

· Prototyping the DiSSCo Digital Architecture and creating a demonstrator (Cardiff, SGN - Senckenberg Gesellschaft für Naturforschung, Naturalis, RBINS - Royal Belgian Institute of Natural Sciences, UTARTU - University of Tartu, RBGE - Royal Botanic Garden Edinburgh) (M1 - M18)

· Demonstration of Machine Learning algorithms integration to serve use cases for knowledge discovery and annotation of digital specimen data. To prototype, demonstrate and describe how to integrate these with the DiSSCo Knowledge Graph to serve data relationship discovery and exploration, semantic interoperability, and data governance (SGN, Cardiff, Naturalis) (M18 - M36).

· Workflows and technical instructions that enable collaborative FAIR data curation and annotation, including machine-added annotations/metadata. Working in close collaboration with Task 5.2, this activity will 1) refine the use cases for bio-/geodata annotations (including scholarly and citizen science applications), 2) explore technical solutions currently implemented across DiSSCo facilities; and 3) measure their compliance with the W3C annotations standards and FAIR principles, with a particular focus on interoperability. This builds on the data standards activities in

ICEDIG, SYNTHESYS+ and the MOBILISE COST (Mobilising Data, Policies and Experts in Scientific Collections) Action (Table 4 – objs. d, g, h, m, o) (M18- M30).

Timeline and Summary Recommendations

Table 12. Implementation of DS Architecture Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing														
Pri	**Rec**		**2023**				**2024**				**2025**				**2026**			
Typ	**No**	**Recommendation**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Tr	043	(Before 2024): **Finalisation of a first release of the openDS specification** (see section "Discussion and Conclusions" for details). Immediate goal is a first release within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
Tr	044	(Before 2024): **Develop a coherent digitisation strategy binding for DiSSCo partners** including set up of infrastructures like sufficient storage capacities and the implementation of specifications (openDS, MIDS) to enable reliable feed into DiSSCo's core data pipelines.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
H	045	**Transfer pilot plans of DiSSCo's service hub (e.g. repo[16c], data pipelines) detailed in D6.2 Implementation and construction plan to full production facilities** (Leeflang 2022[17f]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
H	046	**Develop, coordinate and operate a distributed, cross-institutional service support network to facilitate the local implementation of necessary technical adjustments linked to the Digital Specimen architecture.**	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
H	047	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation **[H] (concept: "UCAS - Unified Curation and**	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

		Annotation System") and annotation [M-L] (prototypes Specimen Data Refinery Hardisty 2022[18g], Machine Learning as a Service (MLaaS) pilot Grieb 2021[10f]).				117												
M	048	**Situate DiSSCo as the major driver for the transformation of NSCs and other biodiversity data providers into integrated hubs for geo- and biodiversity data as part of European and global hyper infrastructures (EOSC, Gaia-X, GOSC) leveraging concepts like the FAIR Digital Twins** (Schultes 2022[45c]). This implies that the current technical focus of DiSSCo be widened by transferring and extending the established methods from collection-based research to other fields of geo- and biodiversity research (e.g. macroecology, biologging). Specifically, this means active provision and support of a FAIR ecosystem of services for the geo*- and bio*-communities involving infrastructures for modelling of new data types (beyond the current focus of the DS), hosting of registries for profiles and operations of those types and provision of service infrastructure for fundamental requirements like PIDs and AAI (cmp. section 5.3. in Wittenburg 2022[46c] for the high-level of FDO-compliance achieved by DiSSCo services).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
M-L	049	**Extend the scale of useful prototype properties of the DiSSCo's Digital Specimen architecture and transfer these into the production system.** Special attention will be given to services for distributed data maintenance and curation [H] (concept: "UCAS - Unified Curation and Annotation System") and **annotation [M-L]** (prototypes Specimen Data Refinery Hardisty 2022[18h], Machine	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

		Learning as a Service (MLaaS) pilot Grieb 2021).				18												
L	050	**Provide a discovery service to cluster/match types of Digital Specimens (and other biodiversity Digital Objects) against requirements for scientific data**. Objective is to enable the determination of semantic similarity of DS and the suitability of the encapsulated data by providing (or ranking) of appropriate Digital Objects using DS attributes transformed to vector representations/embeddings (Le 2014[55c], Devlin 2019[56c], Wittenburg 2022 in sub.)	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

Discussion and Conclusions

· The specification for open Digital Specimens (openDS) is highly aligned with key topics of DiSSCo's e-service infrastructure: OpenDS details the implementation of Digital Specimens as FAIR Digital Objects (FDOs) and acts in this way as a link to the wider framework of Digital Object Architecture (DOA) (Islam 2021[85a]). Semantic data models expressed in openDS are also essential components in the context of the Minimum Information about a Digital Specimen (MIDS[86]) and the permits/loans standard vocabulary for ELViS (European Loans and Visit System)/DiSSCo (SYNTHESIS+ D4.1, NA3.3).

· The Digital Specimen concept provides unified (abstract) views on data originating from different sources by linking to remote information artefacts like sequences stored in an International Nucleotide Sequence Database Collaboration (INSDC)[87a] - International Nucleotide Sequence Database Collaboration or occurrence data in GBIF (Global Biodiversity Information Facility)[88a]. Consequently, the DS Arch is oriented towards the storage of metadata. It uses links to point to the original data, which is to a large extent hosted in remote repositories. (cp. section "DiSSCo Data Storage Container'' in Leeflang 2022[17g]). It emerged clearly from the discussions between DPP WP4, DPP WP6 and DiSSCo's Technical Team that an integrated modular storage concept is required that defines the data streams from local digitisation streams to the DS architecture and is scaled and parameterised to store the large amounts of image data from digitisation pipelines (a quantitative estimate for botanical data is provided in Borsch 2020[89]). A drafting process on such a document started.

Method

DiSSCo's key element to achieve FAIR is the Digital Specimen (DS) which encapsulates and persistently links to all relevant information artefacts about the physical specimen. Examples for such artefacts are sequences stored in the INSDC[87b] databases, occurrence data in GBIF[88b] or traits and ecological interactions in EOL-TraitBank[90](Encyclopedia of Life-Trait Bank). The DS data model serves as principal Digital Object type for geo- and biodiversity objects and embeds in this way collection data in the wider framework of FAIR Digital Objects (Islam 2021[85b]). DS are (like any FDO type) integrated in a FAIR ecosystem of key data services like persistent identifier registration (PID) and terminology services (Collins 2018[91]), registries for type profiles and operations. This enables the self-contained processing of data for software agents (machine-actionability) and repositories for the objects themselves (DiSSCo repo[92]).

To model the DS and other FDO types for Geo- and Biodiversity (including i.a. annotation and machine observation objects), the open Digital Specimen specification (openDS[93]) and a corresponding framework for collaborative modelling (DiSSCo Modelling Framework, DMF[94a]; Fichtmueller 2021[95]) were developed.

Main Methodology (iterative development based on prototyping);

· From the start of DPP, a regular coordination meeting ("openDS breakout group") was established with key developers and modellers of the DiSSCo-linked projects in concern to organise the continuous specification of openDS. The group currently meets bi-weekly.

· Agreed-on OpenDS elements are checked and expanded (e.g. assignment of additional attributes or subclassing) with regard to requirements of the technical architecture by the development team at Naturalis and implemented in the DiSSCo Modelling Framework (DMF[94b]) by ontologists from (Senckenberg Gesellschaft für Naturforschung (SGN) & Botanischer Garten und Botanisches Museum Berlin (BGBM).

· Extension of core services are implemented by the development team at Naturalis accordingly (see Leeflang 2022[17h]), additional enrichment services, e.g. based on Machine-Learning by the team at Senckenberg (Grieb 2021[10g]). Information and results gathered from prototyping are passed back to the openDS breakout group.

4.3.2 Interoperability (Institutional CMSs)

Collection Management System (CMS) systems interoperability and harmonisation

T6.1 establishes specifications and agreements for local DiSSCo facilities to achieve interoperability of scientific data managed in these systems to transform the currently isolated datasets into one unified European collection, building on experience from established systems for data communication (e.g. GFBio German Federation for Biological Data, GBIF) and software development communities (e.g. DINA Digital Information System for Natural History Data).

The task includes prototyping of Application Programming Interfaces (APIs) that are required to provide bidirectional interfaces for achieving the linkages with the data in the facilities through the DiSSCo Knowledge Graph.

Timeline and Summary Recommendations

Table 13. Interoperability (Institutional CMSs) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
Pri	**Rec**		**2023**				**2024**				**2025**				**2026**			
Typ	**No**	**Recommendation**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
H	051	**Install a local dedicated CMS person (developer) responsible for development to implement specific response procedures triggered by new event types.** For small to medium sized museums, we recommend hiring a developer at least for the first three months or for the period agreed upon between the institution and the professional to do the implementation requested by this event-driven API on the local CMS.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
H	052	**Install a local CMS technical team for ensuring compatibility** while both the CMSs and DiSSCo RI evolve.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
H	053	**Provide technical maintenance services and basic software support.**	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

			P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
H	054	**Further, develop, coordinate and operate a cross-institutional DiSSCo CMS support network (DiCSN).** A DiCSN is required to embed local CMS infrastructures of participating institutions in DiSSCo's Digital Specimen architecture. This will be based on the implementation of the API guidelines detailed in DPP D6.1 Harmonisation and migration plan for the integration of CMSs into the coherent DiSSCo Research Infrastructure (Glöecker et al. 2022[19c]) in close alignment with further development of the ingestion data pipelines outlined in DPP D6.2 Implementation and construction plan of the DiSSCo core architecture (Leeflang 2022[17i]). As a major outcome of D6.1, detailed API guidelines for the bi-directional interfaces between local CMSs and the DiSSCo core infrastructure were elaborated. These guidelines could be either implemented as an extension of existing application interfaces to set up DiSSCo specific endpoints in CMSs that are web-based or by provision of a light-weight wrapper, a technique for extracting the data from documents that belong to non-web-based CMSs. To keep the technical integration of the local CMSs synchronised with the ongoing development of the retrieval infrastructure is of major importance to (i) enable continuous data feed into the core architecture (translator services, see D6.2) and to (ii) ensure the reverse data flow back from distributed curation into the local CMSs.																	
M	055	**Develop training and training materials** for the local CMS technical team.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

			P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
M	056	**Further, foster the creation of an ecosystem of interoperable open-source CMS modules and services across the CMS community beyond DiSSCo.** Based on the structures and tools developed within the framework of DiSSCo, the wider community of open CMS developers will be addressed to adopt support for DiSCN-linked interface standards (see D6.1 above) and data models with special emphasis on Digital Extended Specimens and various DES subtypes proposed for openDS (Addink 2022[37e]) as well as for DES/Digitisation-linked standards with special emphasis on MIDS[47c] (MIDS level 3 is closely aligned with openDS/DES.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F
L	057	**Further, evaluate, develop and integrate further functionality** upon request.	P	P	P	P	I	I	I	I	I	I	I	I	F	F	F	F

Discussion and Conclusions

This DiSSCo pilot API and the API guidelines aim to serve as a facilitator for the communication between the collection management systems and the DiSSCo RI. Its functionalities include endpoints for accessing the current events, the possibility of adding new events, and the information and capabilities of the CMS. This pilot will also be a key component within the DiSSCo data architecture enabling the infrastructure only to act based on an event produced by the CMS, having collaboration with the Work Package 5 Common Resources & Standards (Task 5.4 Deliverable: DiSSCo Prepare WP 5 – D 5.5 Construction plans for the improvement of technical infrastructure in the areas of geo-collection data and taxonomic services) and work package 6 (Task 6.2 Deliverable: DiSSCo Prepare WP D6.2 Implementation and construction plan of the DiSSCo core architecture[96] and the task 6.4 from SYNTHESYS+: Deliverable 6.5 Specification and requirements for integration with ELViS).

Method

DiSSCo Prepare WP6, Task 6.1, in close collaboration and exchange with other DPP project partners, aggregated the main challenges for achieving the harmonisation and interoperability between the institutional collection management systems (CMSs) and the

DiSSCo research infrastructure (e.g. syncing basic digital record information, diversity of CMSs, evolving and adapting the CMSs[97]).

In view of the diversity and heterogeneity of all these challenges and also the need to unify the communication between the CMSs and the DiSSCo RI, a common specification / guideline for designing an API which supports interacting with the DiSSCo RI was developed. The use of abstraction layers, such as the JavaScript Object Notation-JSON:API specification, OpenAPI Specification and Cloud Events specification, was implemented to provide a solid foundation for CMSs data aggregation and interoperability.

For collecting user stories and understanding the priorities, an event storming workshop for ranking the most common events in CMS, was held by the work package and in sum of the abstraction layers, a Pilot API was developed for harmonisation and communication between the complex systems.

4.3.3 Technical Services Development and Interoperability with DS Arch

Technical interface requirements of the end-user services
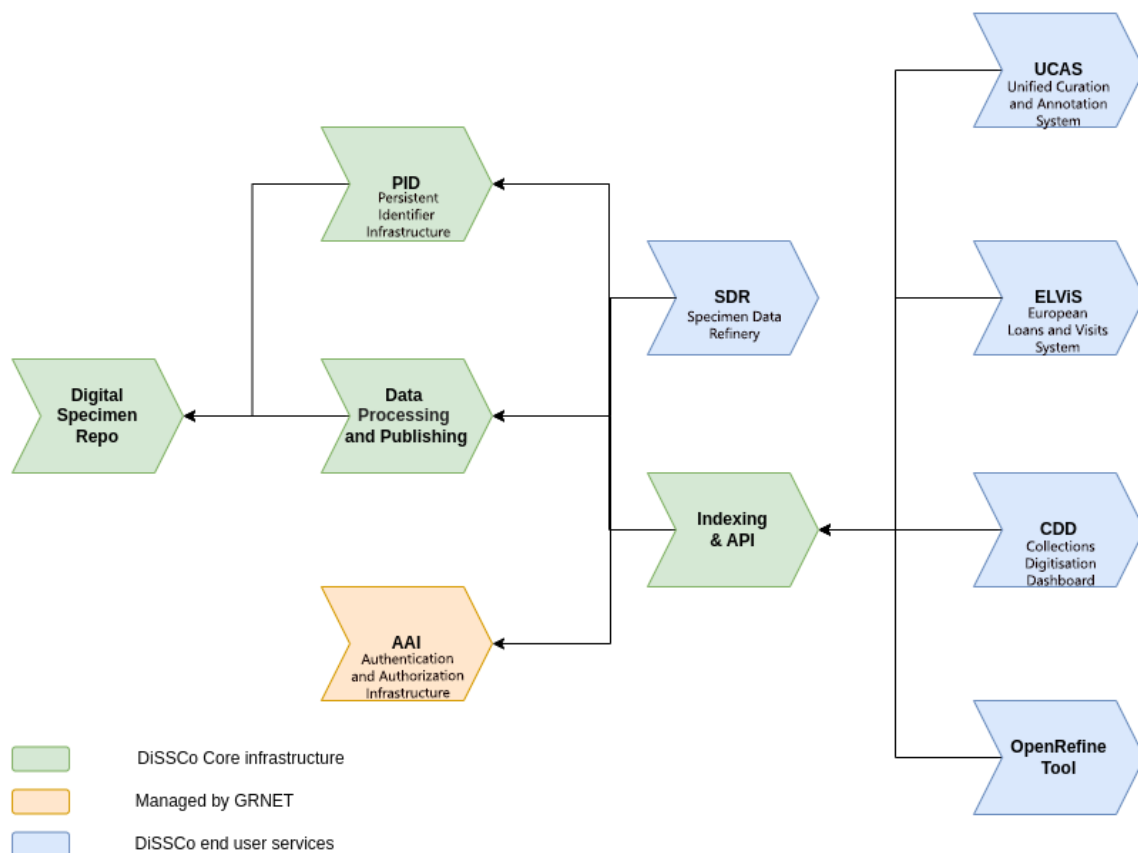
Figure 3. Relationship of DiSSCo core architecture with its services. Arrows indicate dependencies.

This task aims at further defining the collection of technical services and capabilities offered by DiSSCo. Together these services aim to cover the use cases collected from end-users in the ICEDIG project (including individual researchers, other RIs and citizen scientists) and user needs, to be defined in WP 1. In addition, it will describe how the DiSSCo end user services need to interact with the DiSSCo core technical architecture. Interactions with external services like virtual research environments for data aggregation, machine learning, analysis and visualisation (like the Visualization, Analysis and Transformation System tool-VAT within the GFBio Data Federation) tools will also be defined.

The task will make use of other outputs from WP5 & WP6 and results from WP1.

FAIRness maturity requirements for the services will be calculated in the context of WP11 in the ENVRI-FAIR project. The task will also assess the maturity of ELViS, to be developed under SYNTHESYS+ (Table 4 – obj. j), focussing on the adaptations needed for integration with the core DiSSCo architecture.

Timeline and Summary Recommendations

Table 14. Technical Services Development and Interoperability with DS Architecture Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Task Name	Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
				2023				2024				2025				2026			
				1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
[4.3.3 Technical Services Development and Interoperability with DS Arch](#)	H	058	**Find maintainers for the core e-services and plan steps to transform into future DiSSCo service providers.**	P	P	P	I	I	I	I	I	I	F	F	F				
[4.3.3 Technical Services Development and Interoperability with DS Arch](#)	H	059	**Develop the e-services into mature products for operation and integrate them with the DiSSCo FAIRDO infrastructure**	P	P	P	I	I	I	I	I	I	F	F	F				
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	H	062	**Implement Digital Specimen Repo as a component within the DiSSCo Arch.** Tracked in [DD-132](#)[20g].	P	P	P,I	P,I	P,I	F										

				C1	C2	C3	C4	C5	C6	C7
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	H	063	**Adopt the openDS data model.** Tracked in DD-133[23c].		P	P	P,I	P,I	F	F
4.3.3.b Data Processing and Publishing Infrastructure	H	066	**Ensure data processing and publishing functionality** as part of the deployment of the Digital Specimen Repository[25e] (tracked in DD-132[20h])	P	P,I	P,I	P,I	F		
4.3.3.c Persistent Identifier (PID) Infrastructure	H	067	**Build a roadmap for DOI implementation** (from Handle to DOI) Tracked in DD-5[22e].		P	P	P	P,I	P,I	F
4.3.3.c Persistent Identifier (PID) Infrastructure	H	068	**Create a service management plan for the PID service.** Tracked in DD-113[23c].	P	P	P,I	P,I	F		
4.3.3.d Authentication and Authorization Infrastructure (AAI)	H	070	**Finish User Profile development work and deploy it** in the current sandbox. Tracked in DD-13[24c].	P,I	F					
4.3.3.e Indexing and API Service	H	074	**Ensure API functionality** as part of the deployment of the Digital Specimen Repository[25f] (tracked in **DD-132**[20i]).	P	P	P,I	P,I	F		

Component	Priority	#	Description	T1	T2	T3	T4	T5	T6
Components Infrastructure									
4.3.3.e Indexing and API Service Components Infrastructure	H	075	**Provide API documentation.** Tracked in DD-135[26c].			P	P,I	P,I	F
4.3.3.f Unified Curation and Annotation System	H	076	**Finish UCAS Enhancement**. Tracked in DD-26[27e].	P,I	P,I	F			
4.3.3.f Unified Curation and Annotation System	H	077	**Consult users and redesign with UI/UX expert.** Tracked in DD-18[28c].		P	P	I	F	
4.3.3.h European Loans and Visits System (ELViS)	H	083	**Fix and improve current ELViS code base** (ELViS Code Evaluation[29c]). The Epic (DD-16[30g]) (as of Dec 2022) has several tickets in different states of completion.	I	F				
4.3.3.h European Loans and Visits System (ELViS)	H	084	**Align and integrate ELViS with the DiSSCo core infrastructure development** activities **and data management plan.** Also tracked in (DD-16[30h]).	P	I	I	F		

Section			Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4.3.3.i Specimen Data Refinery (SDR)	H	087	**Align SDR with the DiSSCo core infrastructure development** activities. Tracked in DD-6[31c].	P,I	P,I	P,F													
4.3.3.j Collections Digitisation Dashboard (CDD)	H	090	**Automate import of digitised and collection holdings data into CDD** (note: this involves more effort from the collection holdings institutes than the DiSSCo technical team, and will therefore depend on the local technical team, their resources and training).	P	P	P,I	P,I	I,F	I,F	I,F	I,F	I,F	I,F	I,F					
4.3.3.j Collections Digitisation Dashboard (CDD)	H	091	**Align CDD with TDWG standard** (Latimer Core[32e], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). This is related to DD-119[33e] and DD-32[34c].				P	P	P	I	I	I	F	F	F				
4.3.3 Technical Services Development and Interoperability with DS Arch	M	060	**Establish service level agreements with identified service providers**	P	P	P	I	I	I	I	I	I	F	F	F				
4.3.3 Technical Services Development and Interoperability with DS Arch	M	061	**Establish FDO infrastructure compliance requirements for DiSSCo services**	P	P	P	I	I	I	I	I	I	F	F	F				

Section	M	#	Description									
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	M	064	**Align DSR with FDO specifications**.		P	P	P,I	P,I	P,I	F	F	
4.3.3.a Digital Specimen Repository (DSR) Infrastructure	M	065	**Integrate DSR with the DOI ecosystem** (see PID service[98]) Tracked in DD-5[22f] (this has dependency on a few global partners).	P	P	P,I	P,I					
4.3.3.c Persistent Identifier (PID) Infrastructure	M	069	**Ensure FDO compliance** by including Type Registry, FDO Record, and FDO Profile. Tracked in DD-10[48c] (this is also relevant for Digital Specimen Repo[25g]).		P	P	P	P,I	P,I	P,I	P,I	F
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	071	**Implement trust model**[49c] (related to UCAS enhancement work tracked in DD-26[27f]).	P	P	P,I	P,I	F	F			
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	072	**Seek guidance and help from EOSC and EOSC Life** expert resources via GRNet **to build a robust AAI infrastructure**.		P	P,I	P,I, F					

Section	M	ID	Task									
4.3.3.d Authentication and Authorization Infrastructure (AAI)	M	073	**Deploy and use AAI system** for all different DiSSCo services.				P	P	P	P,I	F	F
4.3.3.f Unified Curation and Annotation System	M	078	**Align with openDS data modelling work** for the annotation data model.	P	P	I	F					
4.3.3.f Unified Curation and Annotation System	M	079	**Align with FDO specification.**	P	P	P	I	I	F	F		
4.3.3.f Unified Curation and Annotation System	M	080	**Implement automation and machine annotation to speed up the annotation process** (needs some interaction with the SDR and other AI/ML development activities in the community).		P	P	I	I	F			
4.3.3.g Open Refine Data Management Tool	M	081	**Conduct surveys and consultation with users from organisations that do not have CMS.**		P	P,I	P,I	F				
4.3.3.g Open Refine Data Management Tool	M	082	**Provide a feedback plan to feed further development work.**			P	P	P,I	F			

Ref		No.	Description	1	2	3	4	5	6	7
4.3.3.h European Loans and Visits System (ELViS)	M	085	**Align ELViS with TDWG standard** (Latimer Core[32f], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). Tracked via DD-119[33f].		P	I	I	F	F	
4.3.3.h European Loans and Visits System (ELViS)	M	086	**Ensure coordination and collaboration of distributed DiSSCo core architecture and ELViS code development (potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and ELViS core code base separately. Each of these teams will have different working cultures and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address incompatible workflows. It is also important to identify different milestones aligned with the feature requirements and SRL. Working in such a distributed manner might require training and workshops. The use of agile product roadmap[50c] methodology is highly recommended.	P	P	I	I	I	F	
4.3.3.i Specimen Data Refinery (SDR)	M	088	**Integrate SDR with RO-Crate, Galaxy and FDO specification** (alignment with the BioDT[51c] and BGE[52c] project as RO-Crate might be used there as well).			P	P	P,I	,F	F

			Description																			
4.3.3.i Specimen Data Refinery (SDR)	M	089	**Provide training for local distributed technical teams in distributed ways of working (Potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and SDR core code base separately. Each of these teams will have different working culture and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address this problem. Working in such a distributed manner might require training and workshops.	P	P	P	I	I	F											
4.3.3.j Collections Digitisation Dashboard (CDD)	M	092	**Align CDD with the DiSSCo core infrastructure development** activities. Related to DiSSCo DD-16[30i].				P	P	P	I	I	I	F	F	F					

Discussion and Conclusions

A list of core services has been established, but these are not yet in the desired maturity for operation.

ELViS needs further work to integrate it with DiSSCo architecture; steps have been described in a migration document. An annotations model for the DiSSCo infrastructure is currently being defined and piloted, Specimen Data Refinery (SDR) and other services providing or enhancing data will need to become compatible with this model (based on W3C standard), to be integrated with DiSSCo FAIR Digital Objects (FAIRDO) infrastructure. Several DiSSCo e-services have been piloted or created as beta versions, but agreements for further maintenance, scaling these services up and making them production will need to be set up during the construction phase.

Technical Readiness Levels (TRL) are of limited use to assess the required maturity, therefore we have added the Service Readiness levels (SRL). These SRL levels will indicate what the current readiness of the services is. In each subtask we will define what we see as the target SRL for the service after the Construction phase has been completed.

TRL9
• Actual system proven in operational environment

TRL8
• System complete and qualified

TRL7
• System prototype demonstration in operational environment

TRL6
• Technology demonstrated in relevant environment

TRL5
• Technology validated in relevant environment

TRL4
• Technology validated in lab

TRL3
• Experimental proof of concept

TRL2
• Technology concept formulated

TRL1
• Basic principles observed

SRL8
• The service has been rolled out to its target population

SRL7
• The organisation supporting the service has been adapted as appropriate

SRL6
• Wide-scale adoption: The service (with its technology solution) is adopted by its users and non-users

SRL5
• Evidence of the benefits of the service has been assessed with a statistically significant number of users and non-users

SRL4
• Service prototype developed, tested and validated in lab

SRL3
• Technology identified as an enabler for the desired change

SRL2
• User readiness to change a process or create a new one

SRL1
• Capturing drivers and scoping for change

Figure 4: Service Readiness Levels (SRL) and Technical Readiness Levels (TRL) used to assess the maturity of DiSSCo e-Services. source: EHTEL

Method

A list of core DiSSCo e-services has been defined with earlier discussions in the ICEDIG project and the already planned e-services in SYNTHESYS+ as a starting point. The list has been discussed with the national nodes, been described on the DiSSCo website and in the

form of outreach materials, and current maturity of the e-services was assessed as TRL levels (technical readiness level). Maturity of the ELViS service delivered by Picturae in the SYNTHESYS+ project was assessed and further steps to adjust ELVIS to integrate it with the DiSSCo FAIRDO infrastructure were described in a migration document[99].

Interactions with other DiSSCo services and with external services like the Specimen Data Refinery are being studied through monthly developer meetings. A call for expressions of interest[100] has been created to gather information about the interest of DiSSCo nodes to become a service provider in the future. For this we used SRL levels in addition to TRL levels to define the desired maturity of e-services for operation. Unified Curation and Annotation Service (UCAS) as one of the core e-services is being defined and piloted.

4.3.3.a   Digital Specimen Repository (DSR) Infrastructure

Technical interface requirements of the end-user services
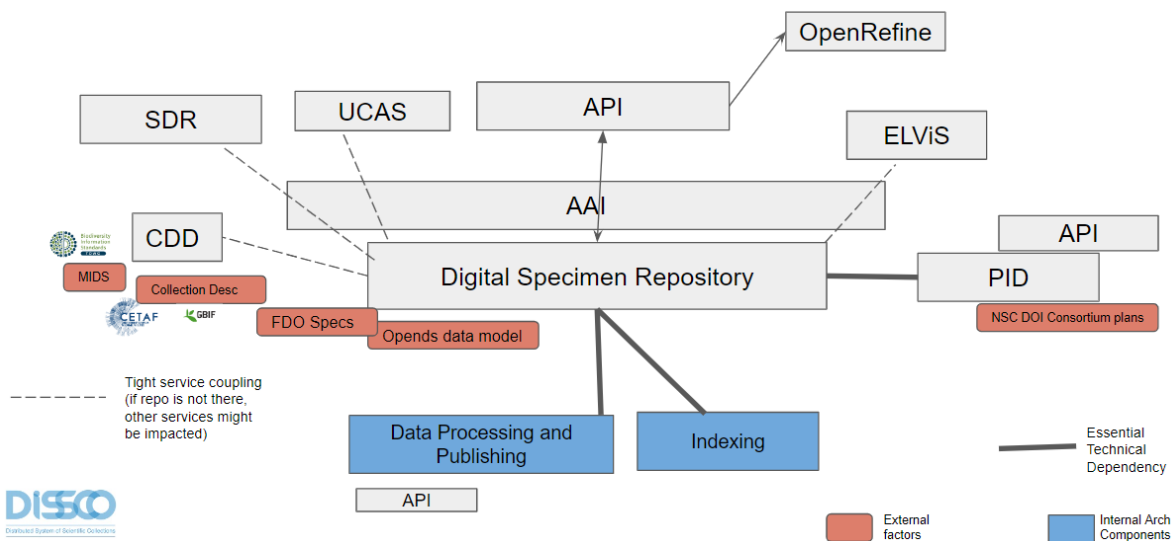
Figure 5. Relationship of the Digital Specimen Repository to other system architecture elements.

Subtask Description   *DPP D6.3*

Digital Specimen Repository is DiSSCo's catalogue of all the Digital Specimens and related FAIR Digital Objects. This is our core data catalogue that provides access to all the information and links related to the digital specimens. The repository will consist of a scalable data storage in which all Digital Specimen data can be stored and retrieved. The ideas and different components described in the DiSSCo core architecture design

deliverables (DPP <u>D6.1</u>[101a]: "Harmonization and migration plan for the integration of Collection Management Systems (CMSs) into the coherent DiSSCo Research Infrastructure" and <u>D6.2</u>[102a]: "Implementation and construction plan of the DiSSCo core architecture") will provide the foundation of this repository.

The Infrastructure Digital Specimen Repository is currently in <u>Service Readiness Level</u> 3 (SRL3) and the target is SRL8.

Timeline and Summary Recommendations

Table 15. Digital Specimen Repository (DSR) Infrastructure Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	062	**Implement Digital Specimen Repo as a component within the DiSSCo Arch.** Tracked in <u>DD-132</u>[20j].	P	P	P,I	P,I	P,I	F										
H	063	**Adopt the openDS data model.** Tracked in <u>DD-133</u>[21d].		P	P	P,I	P,I	F	F									
M	064	**Align DSR with FDO specifications.**		P	P	P,I	P,I	P,I	F	F								
M	065	**Integrate DSR with the DOI ecosystem** (see <u>PID Infrastructure</u>) Tracked in <u>DD-5</u>[22g] (this has dependency on a few global partners).	P	P	P,I	P,I												

Discussion and Conclusions

We already have a strong technical foundation that is robust, modular and uses established industry open source solutions. Input from openDS data modelling and FDO specification work will be important pieces to ensure FAIR implementation. There are two aspects to this implementation. First, the openDS specification (the work was spearheaded within the DPP WP6) provides input on how the Digital Specimens will look like (for example the data structure, which data elements to use to describe the variety of specimens, which elements should be mandatory, optional). Alignment and mapping with the current Biodiversity Information Standards (TDWG) standards are also discussed here. Second, input from the FDO specification work (spearheaded via the FDO Forum where several DiSSCo technical team members are involved) provides the direction to make Digital Specimens FDO. We need to make sure that the Digital Specimens and the Digital Specimen Repository serve the DiSSCo use cases and at the same time aligns with FDO specifications. Therefore, it is

important to continue the alignment and implementation discussions with the current software development work. As PIDs are essential for both Digital Specimens and FDO, the implementation of the PID Infrastructure and connection with the Digital Object Identifiers (DOI) foundation is also relevant here.

Method

During DiSSCo Prepare a development team has piloted several data storage solutions. By making a pilot implementation of the envisioned architecture we were able to measure if the proposed storage solution would fit the needs of DiSSCo. Based on these pilots a Data Architecture[103] has been written with concrete recommendations regarding the implementation of the Digital Specimen Repository. This data architecture takes into account both the functional requirements coming from the user stories and the non-functional requirements.

Currently, an implementation of the data architecture has been constructed. The Digital Specimen Repository forms the basis of persistent data storage. Without it other components won't be able to function, and it is therefore essential for the further development of all other DiSSCo services. In the overview graphic of DiSSCo dependencies (Fig. 3) it is placed completely to the left, indicating its necessity for all other services.

4.3.3.b   Data Processing and Publishing Infrastructure

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

DiSSCo Data Processing and Publishing Infrastructure receives data from a variety of different sources (such as different CMSs, other data catalogues that contain biological and geological specimens and related information). The Data Processing and Publishing Infrastructure contains two types of functionality, described in DPP D6.2[102b] as the "Data Ingestion container" and the "Data Processing container".

The Data Ingestion container focuses on retrieving and harmonising the data from the different sources. For data ingestion both a pull and a push (through the event-driven architecture described in DPP D6.1[101b]) can be used. Harmonisation is essential in the DiSSCo infrastructure. Without harmonised data the FAIR principles with regard to machine-actionability cannot be implemented. Data will therefore be translated from the local format into the openDS data specification.

The Data Processing container receives the harmonised data and checks and validates the data against a schema (an FDO type). After successful validation this service is dedicated to store the data in the Digital Specimen Repository and ensure that the digital specimen receives a unique PID and becomes available for the different end-user services.

The Infrastructure Data Processing and Publishing is currently at Service Readiness Level 3 (SRL3) and the target is SRL 8.

Timeline and Summary Recommendations

Table 16. Data Processing and Publishing Infrastructure Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

| | | | "P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing | | | | | | | | | | | | | | | |
| Pri | Rec | | 2023 | | | | 2024 | | | | 2025 | | | | 2026 | | | |
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	066	**Ensure data processing and publishing functionality** as part of the deployment of the Digital Specimen Repository[25h] (tracked in DD-132[20k])	P	P,I	P,I	P,I	F											

Discussion and Conclusions

Data Processing and Publishing Infrastructure is essential to get the data in openDS specification. Our current pilot infrastructure is robust and modular which provides a good foundation for future development work. As the data sources are coming from local data catalogues, oftentimes that introduces a variety of data quality and standards issues, for instance, non-conformance with Darwin Core or ABCD (Access to Biological Collection Data) elements, missing data such as collection data or geolocation, or inconsistent reference to collection or institution code. Some of these issues have been discussed in different global venues (example: online consultation[104]: Converging Digital Specimens and Extended Specimens—Towards a global specification for data integration).

Besides being able to handle a multitude of data standards, the Data Processing and Publishing needs to ensure synchronisation between the different data storage solutions which provide the Digital Specimen Repository. Currently, this is a single action in which all storage solutions are updated. However, this complicates the ingestion of new data and reduces the performance. During construction there might be a reevaluation if certain data can be updated at a later moment (eventual consistency).

Method

Our current (as of January 2023) implementation uses open source components such as Postgres, Kafka and Kubernetes to create a modular, scalable implementation. It is able to ingest from different sources, such as Darwin Core Archives or BioCase instances, as well as different formats such as Darwin Core or ABCD. From these sources it generates FAIR Digital Objects with a persistent identifier and structured attributes. These attributes (both in the PID record and digital object record level) provide support for linking, interoperability and large scale machine actionability. The data pipeline now can currently create three different types of FAIR Digital Objects: Digital Specimen, Digital Media, and Annotation.

4.3.3.c   Persistent Identifier (PID) Infrastructure

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

The Persistent Identifier (PID) Infrastructure is an essential service component that provides digital object identifier creation and maintenance. This infrastructure will be used by various DiSSCo services including the Data Processing and Publishing service. The PID Infrastructure provides the mechanism to ensure that identifiers are globally unique, persistent and resolvable.

The PID service is currently in Service Readiness Level 3 (SRL3) and the target is SRL8.

Timeline and Summary Recommendations

Table 17. Persistent Identifier Infrastructure Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	067	**Build a roadmap for DOI implementation** (from Handle to DOI) Tracked in DD-5[22k].		P	P	P	P,I	P,I	F										
H	068	**Create a service management plan for the PID service.** Tracked in DD-113[23d].	P	P	P,I	P,I	F												
M	069	**Ensure FDO compliance** by including Type Registry, FDO Record, and FDO Profile. Tracked in DD-10[48d] (this is also relevant for Digital Specimen Repo[25j]).		P	P	P	P,I	P,I	P,I	P,I	F								

Discussion and Conclusions

The PID Infrastructure is an essential component for the DiSSCo services and FAIR implementation. However, the long term sustainability of the PID Infrastructure and the need for identifiers for various natural science collection entities and related biodiversity data go beyond the use case of DiSSCo. This requires a wider global collaboration. To that end, DiSSCo initiated conversations with a number of organisations in the field of biodiversity to formulate a plan for a DOI Registration Agency (RA). The RA is an entity that provides services such as allocating DOI name prefixes, registering DOI names and maintaining metadata. In order to become an RA, the organisation must be a member of the International DOI Foundation. With a robust foundation and technical implementation, the DiSSCo PID Infrastructure will serve the DiSSCo user community. However, a global outlook and support are important for the wider adoption of PIDs for natural science collections and FAIR implementation. The following points are interlinked and the DiSSCo implementation will provide further support for a global adoption. The points below are listed in order of importance and feasibility:

1. Path towards the DOI RA roadmap, the decision on this is not yet final. A possible implementation might be through Datacite with which we will embark on a pilot in early 2023.

2. Use of the PID Infrastructure in Biodiversity Genomics Europe (BGE[52d]) and in Dutch National Project ARISE[105] (Authoritative and Rapid Identification System for Essential biodiversity information). This implies inclusion of Digital Specimen PID in external databases such as GenBank and Barcode of Life Data System (BOLD).

3. Uptake from the user and journal publishers (to cite these specimens with the DOI).

4. Alignment with openDS data model, FDO Type registry, FDO record etc.

Method

Based on the work in the Biodiversity Community Integrated Knowledge Library (BiCIKL) project we have an architectural design document (BiCIKL D7.1[106]) and a test Handle system[107]. This design work consulted DiSSCo's Data Management Plan[108] and the European Open Science Cloud (EOSC) PID Policy[109] as input which provided a few foundational principles such as FAIR and long term sustainability. Based on those principles and extensive background research and consultation[110], a PID architecture design with Handle and DOI was chosen for the design. The Handle System which is one of the essential components of the PID Infrastructure is a system for assigning, managing, and resolving persistent identifiers. Various DiSSCo services will interact with this system (either via API or through individual service interfaces). For example, the Data Processing

and Publishing Infrastructure will create PID's for Digital Specimens, Digital Media Objects and Annotations.
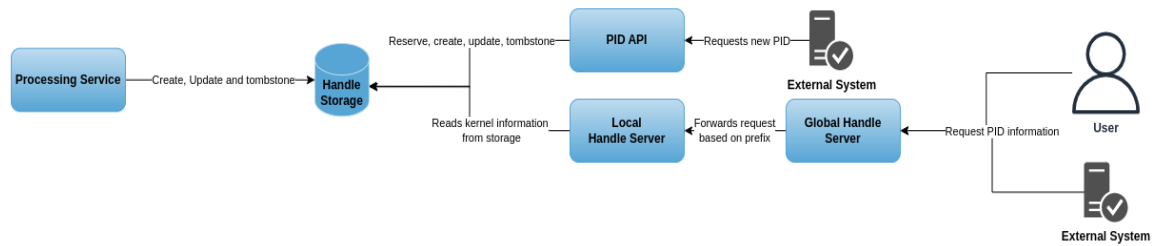
Figure 6. Relationship between the Handle System which assigns, manages and resolves persistent identifiers, and its interfaces.

4.3.3.d Authentication and Authorization Infrastructure (AAI)

Technical interface requirements of the end-user services

DiSSCo will provide a single sign-on Authentication and Authorization Infrastructure (AAI) layer that will interact with different DiSSCo services and Identity Providers (IdP). An AAI infrastructure is essential in securing APIs from the public domain. It ensures that only authenticated and authorised users can perform actions on the data. The design of the AAI component is inspired by the Authentication and Authorisation for Research and Collaboration AARC Blueprint[111] to align with EOSC services. The current demo[112] uses Keycloak as Identity and Access Management (IAM) tooling maintained by the Greek Research and Technology Network (GRNet) and supports different IdP and third party logins (ORCID, Google).

AAI is currently at Service Readiness Level 4 (SRL4). Our target is SRL6.

Timeline and Summary Recommendations

Table 18. Authentication and Authorization Infrastructure Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	070	**Finish User Profile development work and deploy it** in the current sandbox.	P,I	F														

		Tracked in DD-13[24d].																		
M	071	**Implement trust model**[49d] (related to UCAS enhancement work tracked in DD-26[27g]).	P	P	P,I	P,I	F	F											
M	072	**Seek guidance and help from EOSC and EOSC Life** expert resources via GRNet **to build a robust AAI infrastructure.**		P	P,I	P,I ,F													
M	073	**Deploy and use AAI system** for all different DiSSCo services.			P	P	P	P,I	F	F									

Discussion and Conclusions

One possible implementation is a service similar to the Life Science Login[113] (an authentication service from the EOSC Life project[114]; DiSSCo is not part of EOSC Life but the European life science infrastructure for biological information ELIXIR is). This is an exemplary service from an established RI that serves a wide range of users. We could use this as our guidance.

However, during the initial Sandbox pilot phase the following issues were discovered:

1. Not all DiSSCo members are federation-ready and can take advantage of the single-sign on feature. There are workarounds for this so that institute credentials can still be used.

2. We are now working on[115] a customised user profile for each DiSSCo user. These profiles which can provide specific information regarding a DiSSCo user's expertise will be a key for collaboration and community building.

3. To interact with different AAI proxies, each service might need to get permission to use and retrieve attributes of the user. At the moment, we are storing these attributes in a local database which ideally could be fetched from each IdP.

4. User and Expertise profile: we need further work to understand how user and expertise profiles will be used within different DiSSCo services.

5. Trust model: Not AAI specific but related. After authentication and authorisation, we want a dynamic platform for users to interact with the different digital objects, algorithms and services. We also want users to interact with each other to build community around DiSSCo services. This requires user trust and reputation management along with providing enough context to understand the expertise level of the user.

The original impetus for the AAI test and pilot came from the SYNTHESYS+ project to integrate with ELViS. During this project we weren't able to fully integrate ELViS with the AAI. However, we did run a successful pilot where we looked at the possible integration of ELViS and the AAI (see SYNTHESYS deliverable 6.2[116]). This pilot indicated that integration is possible, although it would require some changes to ELViS.

Successful integration of a DiSSCo end user service did happen with both the APIs and UCAS. For these services the AAI was integrated from the start of the project. This generated valuable experience and feedback on the AAI setup.

4.3.3.e   Indexing and API Service Components Infrastructure

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

Application Programming Interfaces (API) are essential for data accessibility and usability. The DiSSCo core infrastructure will expose its data via an API. This API is used both internally, for DiSSCo's own data visualisation, and externally. This enables external users to access the data programmatically but also enables them to build their own data visualisation tools on top of the DiSSCo infrastructure. The API will provide several endpoints, for easy access to the data. It will be built using industry standards such as the JSON:API[117] specification for creating REST (Representational State Transfer) endpoints, JSON-LD[118] for creating linked data and Digital Object Interface Protocol [119](DOIP) to enable connection to other DOA.

For rapid searchability and data retrieval an indexing solution is used. This will ensure that all the data and the fields are fully searchable. Even with large amounts of data, the time to find specific items based on their properties should be minimal.

Timeline and Summary Recommendations

Table 19. Indexing and API Service Components Infrastructure Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
Pri	**Rec**		2023				2024				2025				2026			
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	074	**Ensure API functionality** as part of the deployment of the Digital Specimen	P	P	P,I	P,I	F											

		Repository[25j] (tracked in **DD-132[20l]**).												
H	075	**Provide API documentation.** Tracked in DD-135[26d].			P	P,I	P,I	F						

Discussion and Conclusions

The API and Indexing architecture has been described in DPP deliverable 6.3. The use of industry standards is relatively clear however more experience needs to be gained on how these work together with the FAIR Digital Objects. Both have requirements and in some place these might conflict with each other. For the implementation of DOIP we have yet to run a test implementation. So far the use of DOIP seems limited and the implementations that we have seen take different approaches (for example DOIP over HTTP). Further research is needed on the implementation and how well this works together with other DOIP implementations.

The API and Indexing architecture also touches on security as this is the main entry point into our system for machine-to-machine interactions. This means that this is the place where any malicious activity takes place. During development we make security is therefore on top of everyone's mind. We run regular security checks, both on the code (the checking of Open Web Application Security Project OWASP rules), the build (check for security issues in libraries) and the application. This way we secure our data and our service integrity and provide trustworthy and reliable APIs for our end-users.

Method

Development work has been done in 2022 to build the core foundation for this. There are currently an estimated 40 endpoints in the sandbox environment and this number is growing (documentation available here[120]). These endpoints are used by several of the DiSSCo end-user services such as UCAS and the OpenRefine Data Management Tool. The endpoints have been pointed on top of the Digital Specimen Repository and use the AAI for security. Not all endpoints adhere to the specification described in DPP deliverable 63 (A generalised set of API specifications for interaction with the DiSSCo core architecture). In the coming months we will ensure that the described specification is implemented.

4.3.3.f   Unified Curation and Annotation System (UCAS)

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

Unified Curation and Annotation System (UCAS[121]) is a FAIR annotation and curation platform. It is the main entry point for humans into DiSSCo's data. It provides functionality to search, display, annotate and download the Digital Specimen data. Human experts can add a range of different annotations to Digital Specimens, such as a comment on data quality, a data correction or a linkage to another data source. The annotations are stored as FAIR Digital Objects (with a persistent identifier).

UCAS also enables users to view different versions of the data. This way they can review changes and view the provenance of the object. It also has a view which contains the original data as it was provided by the data provider.

UCAS is now at Service Readiness Level 3 (SRL3). The target is SRL6

Timeline and Summary Recommendations

Table 20. Unified Curation and Annotation System (UCAS) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	076	**Finish UCAS Enhancement**. Tracked in <u>DD-26</u>[27h].	P,I	P,I	F														
H	077	**Consult users and redesign with UI/UX expert.** Tracked in <u>DD-18</u>[28d].		P	P	I	F												
M	078	**Align with openDS data modelling work** for the annotation data model.		P	P	I	F												
M	079	**Align with FDO specification.**		P	P	P	I	I	F	F									
M	080	**Implement automation and machine annotation to speed up the annotation process** (needs some interaction with the <u>SDR</u> and other AI/ML development activities in the community).			P	P	I	I	F											

Discussion and Conclusions

Annotation and Curation on the specimen level data is a highly desired feature. Within TDWG and GBIF, these issues have been discussed a lot over the years. We have taken some of those discussions into consideration during our initial design. However, we need a more targeted user survey and consultation from a DiSSCo user perspective. For instance, we need to better identify who the primary users of this system are (collection managers?

curators? IT dept? data managers? researchers?). Task 2 below has dependencies on how the openDS data modelling work will progress after the end of DiSSCo Prepare (there is no official forum for this yet).

Method

First proof of concept[122] was delivered in Oct 2022 and presented during the 2022 conference of the Taxonomic Databases Working Group (TDWG), today's Biodiversity Information Standards[123]. One full time developer worked on it with the help of the Lead developer. An agile method was used, however not all DiSSCo user stories (collected in 2019) were considered. First feedback was received during TDWG. No significant user surveys were done yet for this tool. Currently, the UCAS system is connected with the Authentication and Authorization Infrastructure (AAI) and the API and Indexing Infrastructure.

At the start of 2023 DiSSCo consulted with an external User Experience/User Interface (UX/UI) expert. They will provide ideas on how to tackle the above-mentioned issues as well as an improved design for the UCAS website. Together with them DiSSCo will evaluate on how to create a pleasurable user experience where information is easily accessible as well as provides functionality for annotating FDOs.

4.3.3.g   Open Refine Data Management Tool

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

OpenRefine is an open source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data. DiSSCo's OpenRefine Data Management Tool will connect with the DiSSCo core architecture and provide the opportunity for pre-upload data cleaning functionality and also help create Digital Specimens. The primary target for this tool is small organisations that don't have a CMS or other structured data catalogue systems. It can also help individual data managers to clean and transform data.

This is currently at Service Readiness Level (SRL3) and the target is SRL5

Timeline and Summary Recommendations

Table 21. Open Refine Data Management Tool Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
M	081	**Conduct surveys and consultation with users from organisations that do not have CMS.**			P	P,I	P,I	F										
M	082	**Provide a feedback plan to feed further development work.**					P	P	P,I	F								

Discussion and Conclusions

The main value for OpenRefine would be for smaller collections and/or organisations without a collection management system. OpenRefine will help with cleaning up data and preparing it for ingestion into DiSSCo. By integrating OpenRefine with DiSSCo, uploading data would mean a press of a button. OpenRefine will also provide functionality to get data back from DiSSCo. It will be able to synchronise the data so that modifications made in DiSSCo will be reflected in the data in OpenRefine. This can help with getting attributes assigned by DiSSCo such as the PID or the MIDS level back into the organisation's data.

OpenRefine could also be interesting for collection managers to use it as their standard data cleanup tool to ensure ongoing data consistency and quality. This requires more investigation and consultations.

Method

Within DiSSCo Prepare WP6, a pilot task[124] was initiated (led by Senckenberg) to understand the feasibility of a tool like OpenRefine with the DiSSCo architecture. In the pilot setup OpenRefine can take in a csv file and send the data to DiSSCo to create Digital Specimens with a PID in our current infrastructure.

The OpenRefine pilot has been built on top of the API and Indexing Infrastructure provided by DiSSCo. OpenRefine uses client credentials to authenticate itself with DiSSCo and after successfully checking its authorisation is able to send data directly into the infrastructure. To synchronise changes back into OpenRefine it uses public endpoints provided by DiSSCo where it can retrieve the latest state of the Digital Specimen.

An important restriction identified during piloting is that OpenRefine has limits to the amount of records it can handle. If the collection grows over the tens of thousands records it might not be the best tool to use. The upload into DiSSCo and the synchronisation is slower as it uses a record-by-record approach.

4.3.3.h   European Loans and Visits System (ELViS)

Technical interface requirements of the end-user services

ELViS[125] is a one-stop shop that provides a unified way to request visits, loans and access through Virtual Access (VA) and Transnational Access (TA) requests. ELViS provides a path for digitisation on demand[126] as a new type of access (whereby users can submit digitisation requests for a collection to be digitised by the holding institutions, to meet a specific need of the research and collections communities), including support for collaborating on proposal/funding call ideas and submission. The request mechanism implemented in ELViS also enables future services for tracking usage metrics, monitoring and reporting and connecting collection usage with research outputs. ELViS Version 1.0 was developed in the SYNTHESYS+[127a] project and released in March 2021. It provided support for the TA and VA calls.

Within DiSSCo we have been looking into Technology Readiness Levels (TRLs) and Service Readiness Levels (SRLs) as a method for estimating the maturity of technologies and services during the preparation, construction, and implementation phase. These levels enable consistent and uniform discussions of maturity and readiness across different types of services and technologies. ELViS is currently at SRL 6 and target (at the end of DiSSCo Construction phase) level is SRL 8.

Timeline and Summary Recommendations

Table 22. European Loans and Visits System Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing																
Pri	**Rec**		**2023**				**2024**				**2025**				**2026**					
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4		
H	083	**Fix and improve current ELViS code base** (ELViS Code Evaluation[29d]). The Epic (DD-16[30j]) (as of Dec 2022) has several tickets in different states of completion.	I	F																
H	084	**Align and integrate ELViS with the DiSSCo core infrastructure development** activities **and data management plan.** Also tracked in (DD-16[30k]).	P	I	I	F														

M	085	**Align ELViS with TDWG standard** (Latimer Core[32g], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). Tracked via DD-119[33g].		P	I	I	F	F										
M	086	**Ensure coordination and collaboration of distributed DiSSCo core architecture and ELViS code development (potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and ELViS core code base separately. Each of these teams will have different working cultures and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address incompatible workflows. It is also important to identify different milestones aligned with the feature requirements and SRL. Working in such a distributed manner might require training and workshops. The use of agile product roadmap[50d] methodology is highly recommended.	P	P	I	I	I	F										

Discussion and Conclusions

As of Nov 2022, ELViS version 1.0 supports TA and VA call functionality with local authentication and roles. Integrating ELViS into the core DiSSCo architecture and development of the loans and visits workflows did not occur during the timeline of the SYNTHESYS+ project. As integration and development of loans and visits workflows are essential functions of ELViS, during DiSSCo's operational phase this should be a high priority. However, for loan functionality to work, the Digital Specimen Repository needs to be in place and ELViS needs to be connected to the DiSSCo core architecture (Recommendation 084 above).

Interdependencies:

- The original scope of SYNTHESYS WP6 Task was dependent on DINA[128] (one of the members of the DINA consortium is part of DiSSCo Prepare) and ColHelper[129] development. The DINA consortium is currently developing an open-source web-based information management system and part of the wider discussion revolves around how different collection and data management systems should interact and integrate with DiSSCo.

- Current ColHelper workflow description and requirements were received during the SYNTHESYS+ work but were not implemented. Discussions about CMS integration also has overlaps DPP Roundtable 2: CMS DS[130].

- ELViS needs to be integrated with the core DiSSCo infrastructure (in particular with DiSSCo's Permanent Identifier Infrastructure and its Digital Specimen Repository). Integration with AAI has been piloted, resulting in SYNTHESYS+ deliverable 6.2[131].

- Personnel and capacity to further develop and maintain DiSSCo e-Services (depends on the responses received from a call for expression of interested parties).

Method

Facilitated via SYNTHESYS+[127b] JRA1[132] online meetings, WP6 members worked with the Picturae[133] development team in defining user stories, requirements, testing, and acceptance. During the SYNTHESYS+ project, the core software deployment was done by Picturae. To facilitate testing and bug reporting public Github[134] was used (Picturae internally used another internal system for their Sprint tracking). In 2019, a survey was conducted to gather user stories and in 2020, an online workflow integration workshop provided further details on system integration. After the deployment of ELViS 1.0, user feedback and a report summary were generated. The components of the current ELViS system include: front end[135] (uses Vue framework) and back end[136] (uses Kotlin[137]). Data storage is done in a Postgres database. For authentication and authorisation, a standalone Keycloak instance was created using a local Postgres database. ORCID identifiers for the users were collected (but not verified) and Global Research Identifier Database IDs for the organisations were stored in an internal database. An external Helpdesk tool was used to facilitate questions during the TA and VA calls.

4.3.3.i   Specimen Data Refinery (SDR)

Technical interface requirements of the end-user services

The Specimen Data Refinery (SDR) provides an easy to deploy, open source, web-based interface. Multiple workflows enable users to enhance existing natural history specimen records (for instance bringing in domain-specific workflows for processing specimen images and extracting text and data from images ref[138]). The SDR uses the Galaxy workflow platform[139] as the basis for managing data analysis, and where possible, using existing Galaxy community tools and approaches.

Timeline and Summary Recommendations

Table 23. Specimen Data Refinery Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
Pri	**Rec**		**2023**				**2024**				**2025**				**2026**			
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	087	**Align SDR with the DiSSCo core infrastructure development** activities. Tracked in DD-6[31d].	P,I	P,I	P,F														
M	088	**Integrate SDR with RO-Crate, Galaxy and FDO specification** (alignment with the BioDT[51d] and BGE[52e] project as RO-Crate might be used there as well).			P	P	P,I	,F	F										
M	089	**Provide training for local distributed technical teams in distributed ways of working (Potential problem):** Distributed teams housed in different institutions will be responsible for the DiSSCo Core architecture and SDR core code base separately. Each of these teams will have different working culture and tools they use for day to day activities. Proper and aligned usage of online tools (such as Slack, Jira, Github etc.) and adoption of DevOps, Sprint/Scrum methods can be a solution to address this problem. Working in such a distributed manner might require training and workshops.	P	P	P	I	I	F											

Discussion and Conclusions

The current SDR Minimum Viable Product (MVP[140]) is a standalone product developed as part of the SYNTHESYS+ project and not integrated with the DiSSCo core infrastructure. Developments regarding the openDS specification have also not yet been implemented in SDR. Several challenges[141] were identified regarding FDO alignment. We are handling these challenges in a two phased approach. In the first phase (from Dec 2022-Mar 2023), the DiSSCo core development team at Naturalis will focus on integrating the SDR code[142] from NHM London in the DiSSCo Core Infrastructure. This will address a few fundamental aspects of what the JSON (JavaScript Object Notation) data objects should look like. Based on this experience, we can provide further specifications on the next development phase. All the SDR related issues are tracked in EPIC DD-6[31e].

Further development has to indicate how various other machine learning/AI activities in different organisations can be aligned with the work in SDR. This will also depend on the responses we receive from the call for Expression of Interest to maintain and develop DiSSCo e-services.

Method

NHM London defined the MVP and led the core development efforts with the University of Manchester. During the development phase, feedback from DPP WP6, openDS breakout team and DiSSCo technical team were provided.

4.3.3.j   Collections Digitisation Dashboard (CDD)

Technical interface requirements of the end-user services

Subtask Description   *DPP D6.3*

The main purpose of the Collection Digitisation Dashboard (CDD) is to make European collections visible, discoverable, easily accessible online, and to highlight institutional contributions, strengths and weaknesses for use by institutions, ministries and other stakeholders. The interactive dashboard[143] visually summarises the digitisation status, content and strengths of collections across the DiSSCo consortium of institutions through a number of visual elements. It displays progress in digitisation and provides summaries and comparisons regarding the number of objects, taxonomic scope, categories of preservation, stratigraphic age, geospatial range, level of digitisation and digital content availability for reuse.

CDD is now at Service Readiness Level 5 (SRL5). The target is SRL8.

Timeline and Summary Recommendations

Table 24. Collections Digitisation Dashboard Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	090	**Automate import of digitised and collection holdings data into CDD** (note: this involves more effort from the collection holdings institutes than the DiSSCo technical team, and will therefore depend on the local technical team, their resources and training).	P	P	P,I	P,I	I,F	I,F	I,F	I,F	I,F	I,F	I,F					
H	091	**Align CDD with TDWG standard** (Latimer Core[32h], a standard for Collection Descriptions) **CETAF Passports, and GBIF Registry** (CETAF and GBIF provide information at the collection description level). This is related to DD-119[33h] and DD-32[34d].				P	P	P	I	I	I	F	F	F				
M	092	**Align CDD with the DiSSCo core infrastructure development** activities. Related to DiSSCo DD-16[30l].				P	P	P	I	I	I	F	F	F				

Discussion and Conclusions

Since the prototype (completed in DiSSCo ICEDIG[1d]), no further development activities have taken place. The prototype focused on manual data entry (using PowerBI). Further discussion is needed to automate the data import and to expand the scope for other organisations to participate. These proposals have the following dependencies:

1. Agreement on standardised ways to describe the collections and digitisation level. This depends on related discussion in the TDWG Latimer Core[144] and Minimum Information Digital Specimen (MIDS[145]) task groups.

2. Each organisation will be in a different stage to provide data resulting in overall data population imbalances.

Method

The <u>current prototype</u>[146] was developed in the SYNTHESYS+ project and released in 2020 with collection information from a limited number of institutions. The prototype is hosted by Natural History Museum, London.

4.3.4 Integration in the Global Technical Landscape

Embedding DiSSCo in the technical landscape

Description

Task 6.4 includes the following subtasks:

Subtask 6.4.1. Embedding DiSSCo in the European and global technical landscape. National efforts and projects (projects (e.g. German GFBio and German Natural Sciences Collections as an Integrated Research Infrastructure DCOLL proposal, US Integrated Digitized Biocollections iDigBio, French Réseau national des collections naturalistes RECOLNAT infrastructure) will be inventoried and a roadmap will be developed to align these with the DiSSCo Construction Masterplan. The Construction plan will also be adapted to international developments in e.g. EOSC, GEOSS, GBIF, EMBL-EBI, ENVRI-RIs, iDigBio, and new recommendations developed in the Research Data Alliance.

Subtask 6.4.2. Representation of DiSSCo in relevant technical committees and standardisation, e.g. bodies, initiatives, organisations or relevant technical working groups for DiSSCo (e.g. TDWG, ELIXIR, RDA).

Subtask 6.4.3. Bring the DiSSCo <u>Data Management Plan</u> (DMP) to maturity including embedding relevant principles and policies with respect to FAIR data policies.

The subtask will assess and update the DMP defined in the ICEDIG project. It will adapt the DMP to the latest developments with respect to FAIR data policies and European rules and best practices for data management.

Measurements will be performed, in collaboration with ENVRI-FAIR and the Go-FAIR (Biodiversity Implementation Network) to evaluate current and anticipated DiSSCo compliance with the FAIR data principles.

Subtask 6.4.4. Enable FAIR data and compliant services, and integration in the EOSC service catalogue.

This subtask will develop and adopt measurements and indicators in collaboration with ENVRI-FAIR and the Go-Fair initiative to measure compliance of data and services with the FAIR data principles. The task will also describe how the DiSSCo services will be added to the ENVRI part of the EOSC catalogue of services.

Timeline and Summary Recommendations

Table 25. Integration in the Global Technical Landscape Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

| | | | "P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing | | | | | | | | | | | | | | | | |
| Pri | Rec | | 2023 | | | | 2024 | | | | 2025 | | | | 2026 | | | |
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	093	**Finalise the first release of the openDS specification** (a technical necessity, see section "Discussion and Conclusions" in CMP T6.4[11c] for details) within Q4 2022.	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
Tr	094	**Ensure close involvement and alignment of DS/DS architecture with developments of Digital Object Architecture, FAIR Digital Objects Forum (FDOF) FDOs including FDO typing, machine actionability, and FDO profiles and attributes.** As noted above, the development of DiSSCo's Digital Specimen concept involving the corresponding service architecture is highly leveraging on Digital Object Architecture. DOA and in particular FDOs as core data models are currently the subject of intensive review and (re-)specification process. The procedure is documented in the FDOF Public Document Management Register[12c]). Of particular importance for the DS/DS architecture are the specifications of FDO typing (important with regard to subclassing DS, see subclasses proposed in Addink 2022 like BotanySpecimen, MycologySpecimen, GeneticResourcePlantSpecimen etc.), Machine-Actionability (the core objective of the FAIR principles to realise a global data ecosystem that can be navigated independently by machines), and FDO PID Profiles and Attributes (which provides essential Kernel Attributes of the FDO when an FDO's PID is resolved, cp. Islam	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

		2020[13d]). It is fundamental for DiSSCo's technical architecture to be fully involved in this process and to ensure a close alignment with delivered results. Currently, members of DiSSCo's technical team are involved in the drafting process of all aforementioned documents with particular relevance for DiSSCo DS architecture.																	
H	095	**Develop the infrastructures for a FAIR service ecosystem leveraging FAIR Digital Objects for the greater range of Biodiversity, Life and Earth System Sciences. Provide the ecosystem of key services for FAIR Digital Objects** (Collins 2018[35c]) like FDO type and operation (for machine agents) registries, services for minting PIDs (cp. Hardisty 2021[36d]), repositories for the FDOs not only for the narrower target group of DiSSCo (NSCs) but also for the greater range of Biodiversity, Life and Earth System Sciences. The fundamental service ecosystem will be supplemented by community services like tools for modelling of new FDO types (for the diversification of DS subtypes see Addink 2022[37f]) or terminology services (TS) for the curation of semantic artifacts, including aforementioned mappings e.g. of DS concepts to classes of the new GBIF data model (a blueprint for a TS embedded in a FDO layer was drafted with participation of DiSSCo as Flexible Semantic Mapping Framework/SEMAF; Broeder 2021[38c]).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?
H-M	096	**Employ the developed FAIR ecosystem in the Bio*, Life and Earth System Sciences to enable advanced modelling, simulation and prediction capabilities.** This implies extending the focus from DiSSCo's core activity	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

		of NSC data management with the key feature of distributed curation during the lifecycle of the DS towards greater activities to support as well (derived) modelling lifecycles. Involved is the creation of new FDO types e.g. for machine-observations, -predictions and biologging plus an even deeper integration of tools and standards for portable workflows like the Common Workflow Language (CWL[39c]) and Research Object Crate (RO-Crate[40c]). This will build upon experiences from the Specimen Data Refinery prototyping in SYNTHESYS+ (Hardisty 2022[18i]).																
M	097	**Assert DiSSCo's role as the major driver of the converging and harmonisation process of the Digital Extended Specimen with respect to establishing a joint digital object approach. Ensure compatibility of data models within approaches** aiming to provide digital object-based information entities like DiSSCo's Digital Specimen and BCoN's Extended Specimen with wider interoperability standards. It is significant that in particular outcomes of the Digital Extended Specimen conversion discussions (Webster 2021[53c]) comply with edge conditions set by overarching FDOt standards (e.g. on FDO Typing).	P	P	P	P	I	I	I	I	I	I	F	F	F?	F?	F?	F?

Discussion and Conclusions

For sustained planning of DiSSCo's DS Arch, oriented to the optimal integration into the global research data management (RDM) landscape, the following principal factors are decisive:

· Consistent alignment of DiSSCo's data model DS with the advancement of GBIF's data model (Wieczorek 2022[147]) to facilitate the highest level of semantic interoperability between DiSSCo and GBIF.

- Assessment and evaluation of the *recommendations for interoperability among* [biodiversity] *infrastructures* worked out as Deliverable 1.2 of the DiSSCo-linked project.

- Biodiversity Community Integrated Knowledge Library (BiCIKL, Meeus 2022 in sub.), e.g. with regard to the aforementioned European Nucleotide Archive ENA/GBIF use case.

- Successful continuation of DiSSCo's participation in the standardisation technological processes of the FAIR Digital Objects Forum[148] which specifies and details the wider framework for DiSSCo's DS concept (for the already high level of compliance of the DS architecture with the concepts of DOA see Table 1 in Wittenburg 2022[46d]).

Development of the Digital Specimen approach as a major driver for a Biodiversity Digital Twin concept within EU initiative Destination Earth (DestinE). DestinE[149] has the aim to develop an accurate digital model of the Earth based on such thematic Digital Twins (DTs) to monitor and simulate natural and human activity. DPP partners (Naturalis, GBIF, Senckenberg) are involved in the Horizon Europe project *Biodiversity Digital Twin for Advanced Modeling, Simulation and Prediction Capabilities* (BioDT, Project ID 101057437) with a particular focus to foster FAIR-compliance of the Biodiversity DT leveraging the FDO approach to make data, software or other research resources like modelling workflows interoperable, reusable and machine-actionable.

Method

DiSSCo's approach to achieve FAIR joint curation, processing and sharing of natural science collection (NSC) data is the implementation of the Digital Twin and the associated services consistently based on FAIR Digital Object (FDO) architecture (Wittenburg 2022[46e]). Both concepts of the FAIR Digital Twin (Addink 2022[150]) and the domain-specific substantiation of the FAIR Digital Object concept in form of the Digital Specimen (DS, which encapsulates and persistently links to relevant information artefacts about the physical specimen) have the objective to embed and integrate DiSSCo's DS architecture in the European and global technical landscape with particular regard to hyper infrastructures like the European Open Science Cloud (de Smedt 2020[151]) or the envisioned global open science cloud[152]. In this context, DiSSCo puts major emphasis on enabling distributed, collaborative, community-based (specimen) data curation and enrichment. Value-added services building upon the DiSSCo Data Enrichment pipeline like machine learning-based annotation and feature extraction[153] detailed in the "Implementation and construction plan of the DiSSCo core architecture" (section DiSSCo Data Enrichment Container; Leeflang 2022[17j]) will be a substantial contribution of DiSSCo to the biodiversity and Earth System Science

infrastructure landscape completing services provided by GBIF as well as by partners in the ENVRI cluster. A very promising pilot study developed under lead of the Meise Botanical Garden involves and combines such infrastructures and services of the European Nucleotide Archive (ENA), GBIF and Wikidata to link sequence data to the voucher specimens and establish consequently bidirectional links from the sequence database to trait, environmental and distribution data. This fabric of bidirectional links was then encapsulated in a DS model which represents in this way an abstracted entity to provide the annotations as permanent references and could in this way enable machine-actionable reuse and repurposing of previously disparate specimen data (Groom 2021[154]).

4.3.5 Geo-collections Mobilisation

Modernising technical infrastructure for science data mobilisation and publication

Description   *DPP D5.5*

The modernisation of key services, especially of services for data currently underdeveloped in the DiSSCo community, is of great importance for the overall improvement of the technical readiness level of the DiSSCo RI.

This aspect of the task focused on construction plans for the improvement of technical infrastructure in the identified key area of geo-collection data, and related geo-classification services. Geo-collection data are highly underrepresented in terms of available services for data mobilisation and publication. Thus, these services need special consideration in terms of their technical capabilities and facility for interoperability, sustainability and resourcing in order to significantly increase DiSSCo technical readiness in the earth science domain.

This task focused initially on the investigation of Geosciences Collection Access Service (GeoCASe) in order to provide a roadmap for integration of this geo-collections aggregator into the overall DiSSCo technical architecture. During the task, conversations with the Earth Sciences community also identified Mindat as a key community resource for authoritative data about geological classifications and localities, and so that platform was added to the review scope of the task.

Timeline and Summary Recommendations

Table 26. Geo-collections Mobilisation Recommendations

NB: ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing			
Pri	Rec		2023	2024	2025	2026

Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	098	In collaboration with TalTech, MfN, the GeoCASe Advisory Board and the CETAF Earth Science Group, **continue to develop the joint vision for the development of GeoCASe,** explore models and opportunities for funding and resourcing GeoCASe development, and assess the overlap with GBIF in the palaeontology domain.	P	I	I	F												
H	099	**Focus initially on publishing data to GeoCASe** through the preparation and presentation of Darwin Core Archives that GeoCASe can ingest.					P	I	F									
H	100	**Incorporate support for Mindat classification and locality identifiers into the openDS specification.**					P	I	F									
M	101	**Explore a potential relationship with Mindat for automated integration between DiSSCo and** mindat.org[54c]. • Collaborate with Mindat to develop a specification and resourcing estimate for integration using DiSSCo's event-driven approach • Investigate potential use cases and appetite for the integration or linkage of DiSSCo Digital Specimen data in the Mindat platform		P	I	I	I											
M	102	**Work closely with GeoCASe on an early proof of concept for event-driven interoperability with an external service** Use the process to: • develop a blueprint and set of minimum requirements for external services to interoperate with DiSSCo using the event-driven architecture; • develop a detailed specification, resource requirements and costings for GeoCASe development to automate ingestion from DiSSCo using the event driven approach.					P	I	I	F								
L	103	**Implement the changes specified in the GeoCASe early proof of concept** if appropriate resources can be made									P	I	I	F				

		available.																

Discussion and Conclusions

DiSSCo's favoured technical model for sustainable interoperability with external services is based on the event publishing approach, originally proposed for integration with institutional collections management systems. An important message that emerged is that this would need investment and collaboration to achieve with GeoCASe and Mindat, as potentially significant development would be required on the side of those platforms as well as DiSSCo in order to ready them for interoperability according to this model, as well as the technical scalability and sustainability that would be needed to meet DiSSCo requirements.

While there are more pragmatic approaches to integration that can be explored in the short term, further exploration of the longer-term relationships and opportunities for external investment form a key part of the task's recommendations. Accordingly, two high-level recommendations for the integration of geo-collections and geo-classification services arose from this task:

1. Apply a phased and pragmatic technical approach in working towards interoperability with external services.

   An event-based approach represents an efficient and scalable method for interoperability in the long term, but is also dependent on third parties prioritising the required development and having resources available, which may not be easy to achieve without investment of funds and/or technical resources from DiSSCo for that development.

   In the shorter term, a more effective approach may be for the DiSSCo development team to continue to pilot more bespoke integrations with GeoCASe and Mindat. These quicker proofs of concept might also help to demonstrate the benefits of interoperability and strengthen the case for funding from external sources to work towards the more robust event-based architecture.

2. Investigate strategic opportunities to address the funding and resource deficit in the geo-collections and geo-classification services.

   There is a clear and critical resourcing issue for GeoCASe at present, with a very small amount of developer time currently available that is due to expire at the end of 2022. In the current situation, GeoCASe will have little if any capacity to implement any recommendations to meet DiSSCo integration requirements, and there is also a risk to the ongoing stability, maintenance, and support of the platform. DiSSCo may, by working in collaboration with the GeoCASe governance bodies, be able to help in

addressing these issues and laying the foundations for a sustainable integration with a stable GeoCASe platform.

Mindat also appears to have a dependency on limited developer resources, which may restrict the platform's ability to deliver the API development and server scalability that would be necessary to support interoperation with DiSSCo at scale. Although initial conversations are yet to be held, there may be opportunities for DiSSCo to work with Mindat to help to progress its development roadmap in this area, if this aligns with Mindat's strategic objectives.

Method

Reviews of the two platforms were carried out through a combination of desk-based research, document review and, in the case of GeoCASe, direct input from developers and representatives of the platforms. User requirements were extracted from earlier DiSSCo outputs, in particular the user stories analysed in Task 1.2, and further detail was added to the picture by running an event-storming workshop, which engaged wider members within the community to help to identify and prioritise those events in external systems that would need to be reflected in the data within the DiSSCo architecture. Technical requirements from architectural design and pilot activities carried out by the DiSSCo developer team were also incorporated, with a focus on an exploration and test implementation of the Biodiversity Collection Access Service (BioCASe) Provider Software for the potential provision of data to GeoCASe[155]. From these activities, it was possible to make a gap analysis between the technical capabilities and resource capacity of the reviewed platforms on one hand, and DiSSCo's requirements and technical roadmap on the other, which in turn formed the basis for a set of recommendations for DiSSCo Construct.

05 Governance, Strategy and Stakeholders

5.1 Overview
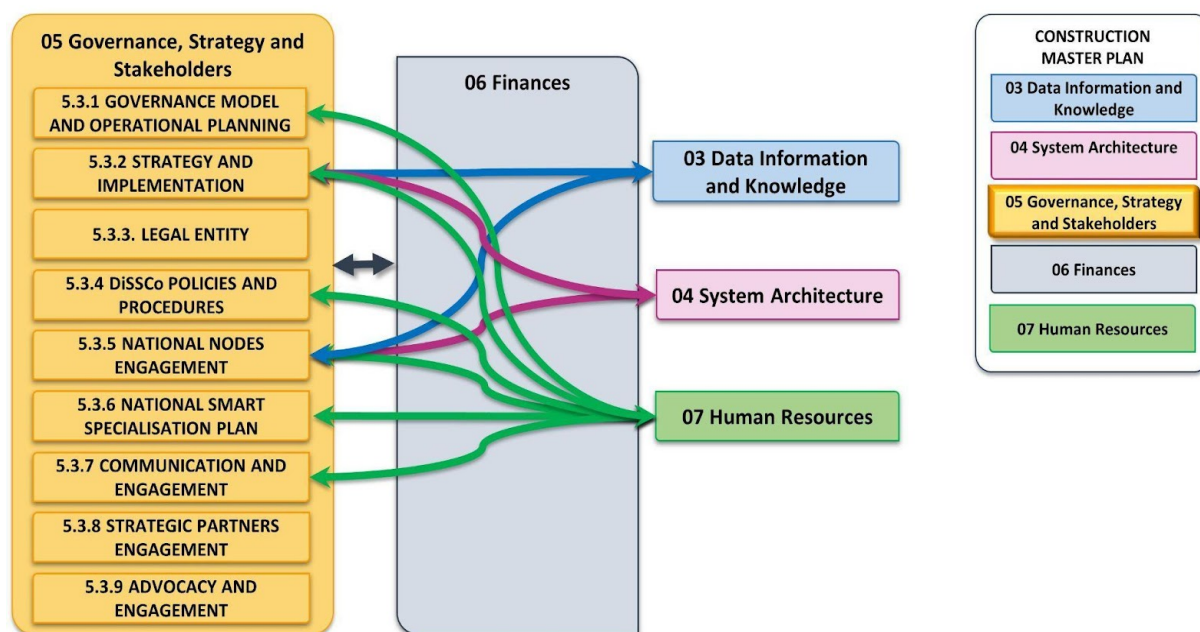
Figure 7. Synoptic graph showing the dependencies (arrows) between tasks in this chapter with other tasks or chapters (systems) in DiSSCo Construct.

This chapter aims to provide complete information about

- the different activities under the business domain carried out during the preparatory phase of DiSSCo;

- the methodology guaranteeing a structured, transparent, inclusive decision process;

- the activities meant to develop during the implementation phase;

- a list of recommendations to guide the implementation phase towards the operation of DiSSCo.

The business pillar of DiSSCo comprises a wide number of institutions across DiSSCo partners and advisory bodies working in:

- governance

- strategic stakeholder engagement

- communication

- specialisation plan

The objective during DiSSCo preparatory phase was to achieve the right maturity level (Implementation Readiness Level, IRL) in governance to guarantee a successful implementation of DiSSCo.

That meant being able to decide on the type of <u>legal entity</u> DiSSCo would adopt in the future, its <u>governance</u> model, <u>strategy</u> as well as its business framework including its <u>contribution model</u>. Furthermore, pursuing an economy of scale, the RI aimed to set up a <u>specialisation plan</u> and a number of tools for the <u>alignment of institutional policies with DiSSCo common policies</u>. Both allow for well-informed decisions about the prioritisation of objectives at national and DiSSCo levels.

To do that, DiSSCo established a methodology that has included main stakeholders along the entire decision-making process (from drafting to decision):

- Communication has been essential to guarantee we share the same language and approach (communication tools) when advocating for DiSSCo or even working together (Binnacle).

- National Nodes (hereby NNs) regular meetings kept DiSSCo partners informed and enabled them to be ready to support the RI whenever necessary.

- The <u>advocacy and engagement strategy</u> structured necessary steps towards the constitution of the Funders Forum (FF) and funders' engagement. The participation of NNs was essential for the FF advisory body constitution, among other success stories.

- The European Research Infrastructure Consortium (ERIC) roadmap in which the NNs and the FF contribute meaningfully has <u>structured a process towards the constitution of DiSSCo's legal entity</u> and guided the work across projects to complementarily reach the objective.

- The extraordinary work that teams did across the wide portfolio of DiSSCo during its preparatory phase deserves also to be mentioned. The work to identify dependencies and synergies across main topics led to a number of rich discussions around governance and strategic partnership, contribution model and development of core infrastructure services, capacity building and policies alignment or communication and strategy.

The implementation phase allows DiSSCo to focus on the constitution of the ERIC and the development of main policies and regulations. The methodology continues respecting the

principles of inclusiveness, simplicity, transparency and adaptability. DiSSCo is a community driven RI on which its strength relies.

The work is built on supporting documentation and teams reflecting the high implementation readiness level reached during the preparation phase.

5.2 Timeline and Summary Recommendations

Table 27. Governance, Strategy and Stakeholders Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

				"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing																
	Pri	**Rec**		**2023**				**2024**				**2025**				**2026**			
Task Name	**Typ**	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5.3.1 Governance and Operational Planning	Tr	104	**Statutes and financial provisions approval (ERIC)**	I	F														
5.3.1 Governance and Operational Planning	Tr	105	**Launch Call for Hosting Country and Founding Members (ERIC)**	P	I	I	F												
5.3.1 Governance and Operational Planning	Tr	106	**Prepare Technical and Scientific Descriptions (ERIC)**	P/I	F														
5.3.1 Governance and	Tr	107	**Prepare Step 1 and submit (ERIC)**			I	F												

Operational Planning																								
5.3.1 Governance and Operational Planning	Tr	108	Prepare bylaws and operational rules (ERIC)	P	I	I	F																	
5.3.2 Governance Strategy and Strategic Implementation	Tr	113	**Identify strategic actions for construction phase and first years of operation**. **(H)**	I	F																			
5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1))	Tr	116	**Keep the DiSSCo Community active** (through regular National Nodes meetings and iGAs) **to support their efforts for motivating their national authorities** to have DiSSCo ERIC established.	I	I	I	I	F	F															
5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1)	Tr	117	**Get formal approval of the Statutes and bylaws by the (representatives of the) countries willing to support DiSSCo ERIC** from the beginning.	I	I	I	I	F	F															
5.3.3 Legal Entity (ERIC Statutes,	Tr	118	**Enter and complete the two-step submission procedure to the EU Commission.**	I	I	I	I	F	F															

Scientific and Technical Description, Step 1)																			
5.3.5 National Nodes Engagement	Tr	121	**(H) Continue supporting the node's involvement in the building process.** During the transition phase, CETAF will act as a mechanism to secure involvement of the NNs and obtain their feedback and contributions; During the Construction phase, once the ERIC is in place, the participation of the nodes will be secured with the Nodes Committee, an Advisory Body to the GA. NB: In the timeline only the Transition phase is covered. For the Construction phase, the Nodes committee shall be implemented under the ERIC governance model.	P	I	I	I	F										
5.3.5 National Nodes Engagement	Tr	122	**(M) Continue to launch surveys and compile information at node level for critical aspects of the DiSSCo architecture** and also for other capacity services (as training), as it remains critical for delivering well-grounded reports and driving good evidence-supported conclusions. To be meaningful, the information collected cannot be exclusively from large institutions but should cover the entire community.	P	I	I	I	F										
5.3.5 National Nodes Engagement	Tr	123	**(M) Foster and enlarge participation across countries with periodic updates from the nodes** on the national priorities and the integration of digitisation activities in the national roadmaps.	P	P	P	I	I	I	I	I	I	I	I	F			

Section		№	Description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5.3.5 National Nodes Engagement	Tr	124	**(H) Encourage the adoption at national level of relevant DiSSCo services and developments** such as <u>ELViS</u> and support the population of relevant components with data (such as the Collections Registry as the feeding pipeline to ELViS). **The compilation of Collections Registry data should be finalised before construction and the linkage with aggregators (such as GRScicoll) shall be formalised during the transition phase to avoid duplication and lack of interoperability.**	P	I	I	I	F															
5.3.5 National Nodes Engagement	Tr	125	**(M) Complete the Specialization Plan before 2024** to serve as a vehicle for the CSO (and once constituted, to the ERIC) to better understand participation at national level.	P	I	I	I	F															
5.3.1 Governance and Operational Planning	H	109	**Prepare Step 2 (ERIC)**			P	P	P	P	P	F												
5.3.1 Governance and Operational Planning	H	110	**Constitute ERIC**										F										
5.3.1 Governance and Operational Planning	H	111	**Keep principles of transparency, flexibility and agile decision-making process as pillars for governance** during operation.																	O	O	O	O
5.3.1 Governance and	H	112	**Use an inclusive methodology (Design Thinking) standard for all preparatory works** concerning ERIC application and operational planning.																	O	O	O	O

Operational Planning																				
5.3.2 Governance Strategy and Strategic Implementation	M	114	**Keep principles of transparency, flexibility and agile decision-making process and Design thinking methodology**, as pillars for updating the Strategy.	O	O	O	O	O	O	O	O	O	O	O	O					
5.3.2 Governance Strategy and Strategic Implementation	L	115	**Ensure DiSSCo Strategy implementation is updated regularly** during the transition and construction phases;	O	O	O	O	O	O	O	O	O	O	O	O					
5.3.5 National Nodes Engagement	L	126	**(L) Create internships and/or student grants at institutional level to overcome detected difficulties (in terms of resources availability)** for collecting relevant information from the different members of each node. This action can be coordinated through CETAF.				P	P	I	I	I	I	F						
5.3.4 DiSSCo Policies and Procedures		119	**Further develop the metadata schema of the policy tool to cover the full range of DiSSCo services**, once policy requirements for these services have been determined. (The metadata schema developed in SYNTHESYS+ NA2.1 only covers the policy requirements for ELViS. Further work will be required.) The metadata schema and policy tool are designed to be extensible, and further policy components can be added.																
5.3.4 DiSSCo Policies and Procedures		120	**Improve the institutional summaries and DiSSCo CSO dashboard.** It is hoped that a proof-of-concept summary																

			and dashboard will be created as part of 7.3, using ELViS as an example, however, more development will be required once policy requirements for other DiSSCo Services have been agreed.																	
5.3.6 National Smart Specialisation Plan		127	The Specialisation tool is delivered as a "tested" tool on a sample of institutions from the NH community and Nodes. Although all the NN currently part of DiSSCo have been contacted the current data entered only represent a limited number of data and thus it will require additional efforts to **obtain from all NNs a representative set of data** leading to the analysis of the specialisations present throughout the community of NH institutions.	I	I	I	I	F	F	F								
5.3.6 National Smart Specialisation Plan		128	**Ensure at least two institutions per node provide data.** The Specialisation tool gives access to the characteristics and assets of research institutions and nodes that are part of the community of natural history institutions. It is therefore recommended that at least 2 institutions per node provide their data to be able to have a critical mass of data available to search into and enable the expected analysis of key specialisations. The development of the tool is finished, although it can further evolve and be refined. Its implementation "I" phase is ongoing with enlarged data sets provided from the NH institutions, with a minimum of 2 institutions per NN having contributed to the data set. We consider that an additional year would be needed to reach this level. The Finalisation phase will be reached when the analysis will be applicable to a representative set of institutions of at least 2 institutions per Node.	I	I	I	I	F	F	F								

Section	#	Action																
5.3.6 National Smart Specialisation Plan	129	**Manage and maintain the Specialisation tool** by a representative of the network and community of natural sciences institutions. The information held in the database represents key assets that should be used diligently. This information will also be enriched by further users of the DiSSCo RI that will rely on the FAIR access and GDPR protocols which will ensure a trustworthy repository.	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
5.3.7 Communication and Engagement	130	**Revise and adapt the Communication Strategy according to the next phases of DiSSCo.** Communication work is a non-stop task that necessarily will evolve as DiSSCo develops, and needs to be customised according to the RI's evolution. Therefore, actions in communication need to be maintained and furthermore, increased during the construction phase towards the operation launch of DiSSCo.		P	P	I	F											
5.3.7 Communication and Engagement	131	**Define and implement an evolving schema for communication material** so that users and interested agents can easily understand the status-quo of DiSSCo, the progress made, and the outstanding stages to address;		P	P	I	I	I	I	F								
5.3.7 Communication and Engagement	132	**During Transition further develop and put into practice, the figure of DiSSCo Ambassador,** outlined during the preparatory phase to address strategically the constitution of the DiSSCo ERIC and later on, during the construction phase, the expansion of its membership;		P	P	I	I	I	I	I	I	I	I	F				
5.3.7 Communication	133	**Encourage increased alignment with sister initiatives in Europe and globally to continue the ongoing processes of linking RIs acting in the field,** through either joint			P	I	I	I	I	I	I	I	I	I	I	I	I	F

Section	No.	Recommendation																	
and Engagement		projects such as BiCIKL, or collaborative strategic partnerships																	
5.3.7 Communication and Engagement	134	**Maintain a person fully dedicated to DiSSCo communication tasks** to keep the targeted audiences duly updated and well informed, and to embed these activities also into the development of the DiSSCo ERIC.	P	P	I	I	I	I	F										
5.3.8 Strategic Partners Engagement	135	**Procurement Strategy:** The **MS8.4 milestone** recommends that DiSSCo develops a procurement strategy in order to align long term priorities and objectives with procurement processes. **This strategy should help DiSSCo to develop partnerships, scale up processes, as well as mitigate risks and improve cost efficiency.** This work is continued by T4.4. **Stakeholder Analysis: MS8.6** details DiSSCo stakeholders, including GBIF, Catalogue of Life, GeoCASe, TDWG and International Barcode of Life. **This report informs stakeholder engagement plans and prioritisation.**																	
5.3.9 Advocacy and Engagement	136	**Revise and adapt the Advocacy and Engagement Strategy** according to the next phases of DiSSCo.	P	P	I	I	I	I	I	I	I	I	I	I					

5.3 Task Summaries

5.3.1 Governance and Operational Planning

Refinement of the governance model and operational planning

Description   *DPP D7.1*

DiSSCo Prepare Task 7.1 contributes to the Organisational Implementation Readiness Level (IRL) by producing reference material for a ready–to–implement organisational model, strategy and operation planning for the governance of the future DiSSCo-ERIC (European Research Infrastructure Consortium) .

Its objective was to provide the governmental structure and the strategy for the DiSSCo-ERIC. Furthermore, the leading team plans to continue working on the operational planning during DiSSCo transition phase.

During the preparatory phase the team has achieved a number of outcomes of crucial importance for the organisational readiness of the RI:

-Firstly, the DiSSCo Memorandum of Understanding was updated to collect the actual relations and regulatory framework between the DiSSCo initiative and the institutional partners.

-Secondly, the governance model for the DiSSCo Legal Entity, DiSSCo - ERIC was selected by decision of the interim General Assembly (iGA). The decision came after a consultation process with DiSSCo National Nodes (NNs) and the assessment of DiSSCo advisory bodies (Scientific Advisory Board [SAB], Technical Advisory Board [TAB], and Funders Forum [FF]).

-Thirdly, the DiSSCo-ERIC Statutes and financial provisions (including national contributions) are currently being discussed by the FF after a thorough consultation process with the National Nodes. The FF is the main actor to make decisions as representatives of national authorities and potential ERIC founders.

Outcomes have been part of a collaborative effort with the DiSSCo Coordination and Support Office (DiSSCo CSO) and the team responsible for the DiSSCo's business framework (DPP WP4).

Further progress is built upon the efforts from the specialisation plan (DPP T8.1 National Smart Specialisation Plan) and a well-defined set of services (DPP T6.3 Technical Services

Timeline and Summary Recommendations

Table 28. Governance and Operational Planning Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	104	**Statutes and financial provisions approval (ERIC)**	I	F														
Tr	105	**Launch Call for Hosting Country and Founding Members (ERIC)**	P	I	I	F												
Tr	106	**Prepare Technical and Scientific Descriptions (ERIC)**	P/I	F														
Tr	107	**Prepare Step 1 and submit (ERIC)**			I	F												
Tr	108	**Prepare bylaws and operational rules (ERIC)**	P	I	I	F												
H	109	**Prepare Step 2 (ERIC)**			P	P	P	P	P	F								
H	110	**Constitute ERIC**									F							
H	111	**Keep principles of transparency, flexibility and agile decision-making process as pillars for governance** during operation.													O	O	O	O
H	112	**Use an inclusive methodology (Design Thinking) standard for all preparatory works** concerning ERIC application and operational planning.													O	O	O	O

Discussion and Conclusions

The governance model, strategy and operational planning have dependencies with work on the definition of the legal entity framework and common policies (WP7), financial contributions models (WP4), DiSSCo key stakeholders (WP8, DiSSCo aspirations working group) operation of DiSSCo Funders Forum (WP9), DiSSCo Human Resources Policy (WP2), and Strategic Partnerships (WP8/WP4).

The work has been able to bring insights and experiences to allow the team to analyse and prepare the basis to start discussions under a framework beyond DiSSCo Prepare project, the ERIC roadmap.

The roadmap aims to define tasks and the timeline necessary to successfully constitute DiSSCo-ERIC at the end of 2024. The work done under T7.1 is part of this roadmap.

We can consider that the governance model is finalised. Therefore, the transition phase addresses activities defined in the roadmap towards the constitution of DiSSCo ERIC and the preparation of the operational implementation, it means, further regulations concerning governing bodies and development of main policies.

We can also consider that there is a sound first draft of the Statutes and financial contributions. Results from the discussion with the NNs and the FF support the launch by the CSO, leading the roadmap, of a call for Hosting country and Founding members of the ERIC who negotiate the final version of those documents and submit them in early 2024 (estimation only).

The work regarding the operational planning is foreseen after the constitution of DiSSCo-ERIC. Internal regulations follow the provisions of the ERIC Statutes and consist of, among others, rules of procedures, development of policies, service-level agreements and other contractual agreements. The work will be developed entirely during 2023.

Method

The methodology used for the preparation of the above mentioned outcomes are based on the following main pillars:

1. Collaboration: the work builds upon existing outcomes and current discussions at different DiSSCo Prepare Project (DPP) work packages concerning the definition of the legal entity framework and common policies (WP7), financial contributions models (WP4), DiSSCo key stakeholders (WP8, DiSSCo aspirations working group) operation of DiSSCo Funders Forum (WP9), and Strategic Partnerships (WP8/WP4).

2. The methodology follows the concept of design thinking, a process in 5 steps meant to guarantee the awareness, understanding, commitment and endorsement of relevant stakeholders at the early stages of the decision process (NNs, SAB, TAB and, FF).

    Following this methodology, the expert team (Task 7.1 team) with the support of a legal advisor organised a series of interviews with well-consolidated ERICs and consultation processes with a wide number of DiSSCo partners, building on top of the preparatory analysis and work done internally at task level.

3. DiSSCo community outcomes are systematically approached by the community through consultation processes. High-level recommendations from DiSSCo advisory

bodies (SAB, TAB, FF) are essential in the decision-making process as well, and therefore, the expert team keeps them abreast of developments regularly.

4. The methodology has provided excellent results and actors perceive it as satisfactory and complete, therefore, the leading team plans to improve it for the coming tasks during the transition phase.

Main steps for the preparation of the Governance model:

· In 2020, an expert team under T7.1 was set up. The team was supported by a legal consultant, X-Officio based on the company's previous collaboration with T7.2 in the preparation for the analysis of the possible RI legal entity models and its history of providing sound advice to RIs, mostly ERICs;

· Comparative analysis of existing models. A series of interviews were conducted with well-established ERICs to learn from their experience. Individual interviews were organised with directors or other senior executives representing BBMRI (Biobanking and BioMolecular Resources Research Infrastructure), EPOS (European Plate Observing System), ELI (Extreme Light Infrastructure), DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure). The exercise provided a sound collection of lessons learned from the experience of people directly involved in the early stages of the ERIC legal entity. The research went on to compare the governance models of other initiatives such as GBIF (Global Biodiversity Information Facility), CETAF (Consortium of European Taxonomic Facilities), ELIXIR (European life science infrastructure for biological information) and EOSC (European Open Science Cloud) in an attempt to benchmark other legal frameworks;

· To keep all partners updated on progress, a folder in Google drive was shared with access to all the information. On top of that, the team opened a notebook in Teamwork, to collect the minutes of the meetings and supporting documentation;

· The legal entity assessment methodology (was and) remains flexible and adapts to needs arising during the process to enable proper endorsement of the work by the community, provision of high-level recommendations from the FF and SAB and well-informed discussions at the iGA. To facilitate this input, T7.1 reorganised its timeline to better align the work with the calendar of meetings of the FF and iGA;

· For the preparation of each of the outcomes, a continuous consultation process was organised with DiSSCo National Nodes based on design thinking methodology. Outcomes were also subject to recommendations from the Scientific Advisory Body;

· A total of 5 workshops were held from September to early December. The workshops focused on in-depth analyses of alternative models, and the role of key stakeholders

(DiSSCo Nodes, CETAF, and Other Stakeholders). In parallel, Task 4.3 <u>EU and National Contributions</u> in cooperation with T7.1 held a number of workshops and interviews to define possible alternative national contribution models. The objective was to feed the governance models with an overview of the financial implications;

· Discussions on each outcome together with results from the consultation processes were part of the agenda of the Funders Forum 3rd meeting (February 2022);

· Key contributions and decisions were taken during the 4th meeting of the iGA (March, 24-25, 2022). <u>The iGA approved model 3 as the most suitable model for DiSSCo-ERIC</u>[156]. Model 3 was preferred and selected because it optimally facilitates input by DiSSCo institutions at both the executive level working closely with the Director General, and as an advisory body to the General Assembly.

Main steps for the preparation of the DiSSCo Statutes and financial provisions:

As part of the mandate given by DiSSCo partners to lead developments through the preparatory phase, DiSSCo CSO defined a roadmap towards the constitution of DiSSCo-ERIC. The roadmap was introduced to the National Nodes, the Funders Forum and finally presented by the Funders Forum Chair, Ana de Castro, to the members of the iGA (during the iGA4)[1].

The roadmap led to different lines of work as follows:

· DiSSCo Prepare's expert team (at WP7 Governance) provided an analysis of existing statutes in well-consolidated RIs and formulated a proposal of elements to take into consideration;

· DiSSCo Technical Team produced a preliminary structure of the <u>Technical and Scientific Description</u> due as part of Step 1 of the ERIC application process;

· DiSSCo Prepare expert team (WP4 Business framework) prepared a preliminary technical report on contribution calculation alternatives and the <u>cost book</u>;

---

[1] FF3 meeting (17th February, 2022); iGA4 meeting (24th March, 2022)

- DiSSCo CSO started the preparatory work supported by a legal advisor, based on outcomes from DiSSCo Prepare (WP7) and the existing ERIC regulation[2] and presented it during the Funders Forum 4th meeting (February, 2022). In that meeting, the Funders Forum members decided to create an *ad hoc* writing group with the direct participation of four representatives from the advisory (Netherlands, Italy, France and Belgium);

- The Statutes Writing Group (SWG) met regularly for three months and prepared a first draft of the statutes that launched the consultation process with DiSSCo National Nodes in early October 2022.

5.3.2 Governance Strategy and Strategic Implementation

DiSSCo Strategy

Description   *DPP D7.1*

A key activity in the final phase of the DiSSCo Prepare Project (DPP) towards the transition of DiSSCo to its construction phase is outlining and defining the DiSSCo Strategy. The DiSSCo Strategy is a crucial document that acts as:

- A roadmap for decision making i.e., what do we do and what we do not do?

- External communication material i.e., what is DiSSCo and how to engage with DiSSCo – Who we are.

- Internal engagement document i.e., what has DiSSCo to offer to its community – Why we exist.

The DiSSCo Strategy is not only crucial, but also, an opportunity for community convergence and alignment through an internal and external community consultation process on the strategic priorities of the DiSSCo Research Infrastructure (RI).

Timeline and Summary Recommendations

---

[2] The legal basis supporting the work consists of the Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium[5] (hereby referred to as "the ERIC Regulation"), and the Practical Guidelines for applicants issued by the Directorate-General for Research and Innovation of the European Commission, as well as publications produced by the ERIC Forum (https://www.eric-forum.eu/) and the Organisation for Economic Co-operation and Development (OECD).

Table 29. Governance Strategy and Strategic Implementation Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	113	**Identify strategic actions for construction phase and first years of operation**. **(H)**	I	F														
M	114	**Keep principles of transparency, flexibility and agile decision-making process and Design thinking methodology**, as pillars for updating the Strategy.	O	O	O	O	O	O	O	O	O	O	O	O	O			
L	115	**Ensure DiSSCo Strategy implementation is updated regularly** during the transition and construction phases.	O	O	O	O	O	O	O	O	O	O	O	O	O			

Discussion and Conclusions

Based on the community contributions we have identified four major focal areas that become the programmatic pillars of the DiSSCo Strategy.

The four strategic pillars (or focal areas) include:

    A. Digitisation

    B. Access

    C. Capacity Building

    D. e-Services

DiSSCo Digitisation programme

Goal: Accelerate and lower barriers for digitisation across European Natural History Collections. Strategic objectives:

·   Developing strategies for prioritising digitisation at various levels based on community consensus.

·   Develop open digitisation workflows, training and best practices.

·   Establish & operate regional digitisation centres & centres of excellence.

- Select and implement the use of annotation tool(s) consistent with capabilities of Specimen Data Refinery (SDR) and integrate other developing technologies and resources where appropriate (Artificial Intelligence-AI, robotics, citizen science).

DiSSCo Access programme

Goal: Create, measure and support multimodal standardised routes to access collections.

Strategic objectives:

- Enlarge the scope of ELViS together with a robust access policy and institutional policies.

- Establish an inclusive physical and remote access programme (triage physical access requests based on digital availability of collections).

- Attract users/requests external to our community (e.g. industry, cultural heritage).

DiSSCo Capacity Building programme

Goal: Users and partners are confidently supplying/using data through DiSSCo and are making use of DiSSCo's services.

Strategic objectives:

- Partners understand their level of digital maturity across a range of DiSSCo's services and have targeted areas for improvement.

- Training available within the DiSSCo consortium to build in-house capacity whenever needed.

- Continue to play a leading role in FAIR digital objects standards development and implementation and the requisite education and training of consortium members to ensure that DiSSCo's data and services are FAIR and have impact on research and scientific/societal innovation.

DiSSCo e-Services programme

Goal: Development of e-Services based on user needs defined through engaging with our stakeholders.

Strategic objectives:

- Facilitate a novel community curation model that pulls together expertise from a wider pool of researchers and taxonomists.

- Achieve FAIRness; all data having Persistent identifiers (PIDs) and machine actionable metadata.

- Automate specimen enhancement services, implementing AI technologies to increase the volume and improve the quality of specimen-linked data.

- Make e-Services sustainable through the DiSSCo business model, ensuring human capacity and continuous service management.

Method

The DiSSCo strategy builds upon the previous work done by an *ad-hoc* working group appointed by the DiSSCo interim General Assembly (iGA) to develop a concept for DiSSCo vision and mission, the aspiration refinement task group.

The strategy is conceived as a living document defining the direction to follow during the operation of the RI. However, it is foreseen to be regularly updated if necessary.

Main steps towards the DiSSCo strategy and strategic implementation carried out by the DiSSCo Prepare expert team:

- An expert team was set up to lead the discussions;

- A series of interviews with well-consolidated ERICs to understand the level of detail and content of existing RIs. The outcome showed a great variety of approaches. In general the strategies do not foresee strategic actions;

- Two workshops were organised with the objective to define a strategy based on four main activity pillars for DiSSCo (identified below). The exercise included a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis;

- The expert team prepared a document introducing the main strategic pillars that was presented during the 4th meeting of the Funders Forum (FF) advisory body (June, 2022);

- The document is circulated among DiSSCo stakeholders as part of the consultation process during 2023 and results are discussed during the 6th FF meeting and presented for approval during the next iGA meeting in early 2023.

5.3.3 Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1)

Towards the creation of a legal entity

Description   *DPP D7.2*

In coordination with T7.1, the team has prepared the necessary steps to establish DiSSCo as a legal entity in order for the RI to deliver its services in the most efficient way. The following actions have been performed in sequence:

- Detailed analysis of the legal entity models, including a SWOT analysis, and their suitability for achieving DiSSCo objectives in agreement with its governance model. The result of the analysis has been presented and discussed with the various national contact entities (cf. WP 8), in order to reach a consensus on the best choice;

- Draft statutes have been prepared, according to the legal entity model that appears to be the most appropriate.

Timeline and Summary Recommendations

Table 30. Legal Entity (ERIC Statutes, Scientific and Technical Description, Step 1) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
Pri	**Rec**		**2023**				**2024**				**2025**				**2026**			
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	116	**Keep the DiSSCo Community active** (through regular National Nodes meetings and iGAs) **to support their efforts for motivating their national authorities** to have DiSSCo ERIC established.	I	I	I	I	F	F										
Tr	117	**Get formal approval of the Statutes and bylaws by the (representatives of the) countries willing to support DiSSCo ERIC** from the beginning.	I	I	I	I	F	F										
Tr	118	**Enter and complete the two-steps submission procedure to the EU Commission.**	I	I	I	I	F	F										

Discussion and Conclusions

The cornerstone of the legal entity, the Statutes, has been discussed during the Funders Forum 5th meeting (FF5). The last discussions before approval take place in early Spring. From that moment, the discussions are handed over to those countries committing to participate in the ERIC, that is, the Host Country and Founding members.

There are strong dependencies with T7.1 and WP4, in order to draft consistent Annexes and Bylaws as part of DiSSCo transition phase.

Method

As part of the mandate given by DiSSCo partners to lead developments through the preparatory phase, DiSSCo Coordination and Support Office (CSO) defined a roadmap towards the constitution of DiSSCo-ERIC. The roadmap was introduced to the National Nodes, the Funders Forum and finally presented by the Funders Forum Chair, Ana de Castro, to the members of the interim General Assembly (during the iGA4)[3]

The roadmap led to different lines of work as follows:

1. DiSSCo Prepare's expert team (at WP7 Governance) provided an analysis of existing statutes in well-consolidated RIs and formulated a proposal of elements to take into consideration. In order to select the most suitable model, the partners:

- identified the criteria and characteristics of the DiSSCo objectives and governance,

- screened other RIs from the same Environment Domain, with the same distributed type and entered on the 2018 Roadmap. But also obtained experiences from other RIs belonging to other domains that have chosen to become a legal entity,

- obtained the support of a legal advisor to go deeper in the legal assets of each entity model available for DiSSCo.

The following six legal entity models:

- International/Intergovernmental Organisation (IO)

- European Research Infrastructure Consortium (ERIC)

- European Economic Interest Grouping (EEIG)

- European Groupings of Territorial Cooperation (EGTC)

- Belgian International Association of the European Heritage Network (AISBL)

- Dutch Stichting

were screened according to 15 points of attention:

1. Is it a separate legal entity?

2. Does it have the capacity to contract with third parties, hire personnel, open a bank account, buy equipment, sue and be sued, *etc.*?

3. Can institutions from EU/EEA Member States be members?

4. Can institutions from non-EU/EEA Member States be members?

5. Can EU/EEA governments (ministries) be members?

6. Can non-EU government (ministries) be members?

---

[3] FF3 meeting (17th February, 2022); iGA4 meeting (24th March, 2022)

7. Is the establishment process simple and fast?

8. Is the legal entity exempted from the EU public procurement directives (Art. 9(1)(b) Directive 2014/24/EU)?

9. Is the legal entity exempted from the common system of value added tax (in the sense of Articles 143(1)(g) and 151(1)(b) of the VAT Directive)?

10. Is the legal entity able to carry out economic activities?

11. No initial capital requirements.

12. Is there a limited liability regime (legal entity, governing body and for members)?

13. Is the governance structure flexible?

14. Is the legal entity able to receive EU and national grants?

15. Is the legal entity able to contract bank loans?

Thanks to this first screening, three models were shortlisted: International Organisation, ERIC and the Belgian AISBL. As International Organisations present more or less the same pros and cons than ERICs but at the price of a much heavier establishment process, the actual final choice was to make between ERIC and AISBL.

The final round, based on specific practical criteria:

	AISBL	ERIC
Level of integration between the DiSSCo legal entity and the National Nodes	+	++
Location of the statutory seat	+	++
Language to be used for the official documents	+	++
Membership	Institutions, individuals	Countries, IOs
Participation by CETAF	++	+
"Branding" and network possibilities	+	++

Procurement and VAT considerations	+/-	++
Tailor-made governance structure	+	++
Securing funding	+/-	++

led to the conclusion that the ERIC legal entity model was the most suitable for establishing DiSSCo RI as a sustainable organisation.

2. DiSSCo Technical Team produced a preliminary structure of the <u>Technical and Scientific Description</u> due as part of Step 1 of the ERIC application process;

3. DiSSCo Prepare expert team (WP4 Business framework) prepared a preliminary technical report on contribution calculation alternatives and the <u>Cost Book</u>;

4. DiSSCo CSO started the preparatory work supported by a legal advisor, based on outcomes from DiSSCo Prepare (WP7) and the existing ERIC regulation[4] and presented it during the Funders Forum 4th meeting (February, 2022). In that meeting, the Funders Forum members decided to create an *ad hoc* writing group with the direct participation of four representatives from the advisory (Netherlands, Italy, France and Belgium);

5. The Statutes Writing Group (SWG) met regularly for three months and prepared a first draft of the statutes that launched the consultation process with DiSSCo National Nodes in early October 2022.

- · The Statutes and financial provisions were discussed during the Funders Forum 5th meeting (14-15 December 2022). In Spring the Funders meet again to finalise the last version of the Statutes to hand over to the representatives of the Host Country and Founding members (FF6).

---

[4] The legal basis supporting the work consists of the Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium[5] (hereby referred to as "the ERIC Regulation"), and the Practical Guidelines for applicants issued by the Directorate-General for Research and Innovation of the European Commission, as well as publications produced by the ERIC Forum (<u>https://www.eric-forum.eu/</u>) and the Organisation for Economic Co-operation and Development (OECD).

5.3.4 DiSSCo Policies and Procedures

Develop and establish DiSSCo policie

Description

The main objective of this task is to provide all institutions in the DiSSCo RI with a "direction map", identifying a series of targets and possible pathways to reach them, in order to have their policies align with a common set of principles and rules that support DiSSCo services. Assessment tools are also provided to help institutions identify the most crucial issues. This task looks at the key concepts underlying the following policies: (i) Physical access policies (which are now rather homogeneous and mature thanks to the SYNTHESYS [Synthesis of systematic resources] projects) (ii) Data policies and digital access policies (access to and use of the products of the digitisation processes); and (iii) Collection curation policies.

With the completion of this task DiSSCo Prepare provides:

- A map of legislation, policies and policy components that are relevant to physical/digital access and digital collections. This can act as a policy self-assessment tool for DiSSCo facilities and help to identify gaps in and develop institutional policy documentation, reducing the requirement for local resource and expertise;

- Improvement of discoverability and interoperability of policy documentation across DiSSCo facilities, so institutions can reuse relevant parts of existing documents for drafting policy documentation;

- Mapping of common DiSSCo policies into the framework to support the alignment process. The different regulatory levels are addressed by the map, through several case studies. This work is largely based on and benefits from the ongoing efforts in ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage)(Table 4 – obj. e) and from the contribution of the various national contact entities (cf. WP 8 Stakeholder Engagement and Communication). Interaction and coordination with the "FAIR Policy Working Group" set up in the frame of the H2020 ENVRI FAIR (Table 4 – obj. w) are also established.

Timeline and Summary Recommendations

Table 31. DiSSCo Policies and Procedures Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
Pri	**Rec**		2023				2024				2025				2026				
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

119	**Further develop the metadata schema of the policy tool to cover the full range of DiSSCo services**, once policy requirements for these services have been determined. (The metadata schema developed in SYNTHESYS+ NA2.1 only covers the policy requirements for ELViS. Further work will be required.) The metadata schema and policy tool are designed to be extensible, and further policy components can be added.																			
120	**Improve the institutional summaries and DiSSCo CSO dashboard.** It is hoped that a proof-of-concept summary and dashboard will be created as part of 7.3, using ELViS as an example, however, more development will be required once policy requirements for other DiSSCo Services have been agreed.																			

Discussion and Conclusions

The final deliverable was due in July 2022, and the policy tool should be at the Technical Readiness Level TRL5-6 by the end of the DiSSCo Prepare Project (DPP) project.

This task links closely with T3.1 (Capacity Enhancement), as the tool is being developed to also support the digital maturity self-assessment tool. The modularity of the tool should enable it to be integrated with other self-assessment services in future.

Method

The self-assessment policy tool is currently under development. User stories were gathered through contributions from task partners and interviews with potential users of the tool. MS7.5[157] contains the design blueprint for the policy tool, including functional and non-functional requirements. These were ranked using MoSCoW ('Must have', 'Should have', 'Could have', 'Won't have').

The metadata schema, which forms the backbone of the policy tool, was developed as part of SYNTHESYS+ NA2.1, and contains policy requirements relevant to the ELViS (European Loans and Visit System) e-service. The metadata schema forms the basis of the questions asked by the self-assessment, and enables the results to be presented as an institutional summary to the user and as a dashboard for the DiSSCo CSO.

Development of the policy tool is underway, with the final deliverable due in July 2022.

5.3.5 National Nodes Engagement

DiSSCo national nodes engagement

This task aims to engage with the broadest DiSSCo community and involve them in achieving the major objectives settled throughout the project in which the participation of the DiSSCo National Nodes (NNs) is considered instrumental to acknowledge the differences across countries, to collect the necessary backgrounds and to streamline the exchange of relevant information from the DiSSCo actors to the interested agents in the surrounding landscape, including policy-makers and governmental representatives. Save for those countries in which the national consortium has not been formed yet, each participating country has designated an institution that represents the national node, through a Node representative. In the former, as it is the case of Belgium, UK and Germany, there are different national institutions represented at the NNs group and attending the NNs meetings

Timeline and Summary Recommendations

Table 32. National Nodes Engagement Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	121	**(H) Continue supporting the node's involvement in the building process.** During the transition phase, CETAF will act as a mechanism to secure involvement of the NNs and obtain their feedback and contributions; During the Construction phase, once the ERIC is in place, the participation of the nodes will be secured with the Nodes Committee, an Advisory Body to the GA. Please note: In the timeline only the Transition phase is covered. For the Construction phase, the Nodes committee shall be implemented under the ERIC governance model.	P	I	I	I	F											

Type	No	Action																	
Tr	122	**(M) Continue to launch surveys and compile information at node level for critical aspects of the DiSSCo architecture** and also for other capacity services (as training), as it remains critical for delivering well-grounded reports and driving good evidence-supported conclusions. To be meaningful, the information collected cannot be exclusively from large institutions but should cover the entire community.	P	I	I	I	F												
Tr	123	**(M) Foster and enlarge participation across countries with periodic updates from the nodes** on the national priorities and the integration of digitisation activities in the national roadmaps.	P	P	P	I	I	I	I	I	I	I	I	F					
Tr	124	**(H) Encourage the adoption at national level of relevant DiSSCo services and developments** such as ELViS and support the population of relevant components with data (such as the Collections Registry as the feeding pipeline to ELViS). **The compilation of Collections Registry data should be finalised before construction and the linkage with aggregators (such as GRScicoll) shall be formalised during the transition phase to avoid duplication and lack of interoperability.**	P	I	I	I	F												
Tr	125	**(M) Complete the Specialization Plan before 2024** to serve as a vehicle for the CSO (and once constituted, to the ERIC) to better understand participation at national level.	P	I	I	I	F												
L	126	**Create internships and/or student grants at institutional level to overcome detected difficulties (in terms of resources availability)** for collecting relevant information from the different members of each node. This action can be coordinated through CETAF.				P	P	I	I	I	I	F							

*In the timeline only the Transition phase is covered. For the Construction phase, the Nodes committee is implemented under the ERIC governance model.

Discussion and Conclusions

To channel the participation described above, recurrent monthly meetings have confirmed to be a very successful tool. Participation of national nodes every last Thursday of the month (save for summer and Christmas breaks) for a 1 hour meeting has been not lower than 80% and has allowed the two-way communication between the CSO and the Task leader CETAF, and the national nodes. Major updates on the status-quo of the DiSSCo preparation work have been shared, together with major achievements and outcomes (from different WPs) for which the participation has been essential, from the DiSSCo architecture (WP6) to its governance (WP7) and cost (WP4) as well as identification of potential DiSSCo user communities (WP1) and definition of training needs (WP2). On the other hand, different national nodes have used this platform to share their concerns but also successes reached in terms of national roadmaps and strategic priorities in relation to digitisation of collections. Along with these meetings, surveys and calls for information have been launched, promoted and fostered for open consultation on relevant matters.

Together with the participation of national nodes, the constitution of the interim General Assembly (iGA) and the Funders Forum, have been the scenarios where the needs and requirements coming from national participants have matched with the solutions and mechanisms developed under DPP to progressively construct DiSSCo RI.

Several items need to be highlighted as conclusion remarks:

· engagement through direct participation is proven efficient and meaningful;

· involvement from earlier stages in the process and in a recurrent format has facilitated understanding, and acknowledgement of main features of the DiSSCo RI which has driven countries to easily take ownership of the process and its added value to the community, also at national level;

· participation of the nodes representatives is crucial to channel information and involvement;

· a strong node's representation ensures broader reach, better alignment and more consistent implementation.

· compilation of relevant information takes time and requires dedicated resources.

To ensure this model of national engagement does not stop with the finalisation of DPP and before any follow-up project (for the Transition phase) can enter into force (which is planned not earlier than end of year 2023), CETAF has committed to fill this time gap and thus support the nodes engagement during this "interim" time, with the same format and

procedures, i.e. through monthly meetings which intend to be a forum for discussion and exchange of relevant information among DiSSCo nodes. The identical model with CETAF convening and fostering nodes' engagement is kept during the Transition phase since the envisaged budget is not sufficient to allow participation of all nodes as beneficiaries of the DiSSCo Transition project. Furthermore, once the ERIC is in place and its governance structure is consolidated, it is expected that the node participation in DiSSCo construction is channelled through the Nodes Committee, an advisory body to the DiSSCo ERIC GA.

Method

All partners to the DiSSCo Prepare project benefit from 2 PMs to allow them to undertake an active participation in multiple facets and dimensions of the DiSSCo RI development. This assigned time has proven insufficient even though the contributions from nodes have been enormous, much higher than it was originally planned and budgeted. The need of detailed information to be collected at national level to better understand the diversity of scenarios to cover across countries has led to the request of DiSSCo nodes to make thorough research, fill in surveys and compile documentation that has enabled setting the basis for many of the outcomes delivered under other tasks, from the different cost elements to the national priorities, from the development of e-services to the structure of governance of the RI. Moreover, the nodes have been pivotal for raising awareness and activating involvement of national governments in the process of building up DiSSCo and therefore attracting their commitments, political and also financially.

The methodology used to facilitate and activate engagement from the community is based on participation and open consultation of DiSSCo's base constituency. DiSSCo preparation implies processes that cannot be detached from the national representatives that are involved in the operation of the RI. Therefore, collecting their insights from earlier stages was always considered a key driver for success to facilitate endorsement of results and foster national commitments, beyond the participation of the community actors (the collections holders). Under this participatory approach, the high-level strategic points (such as governance structure or operation principles) were presented, discussed and channelled through the national nodes towards the governmental representatives to get feedback that would ensure that the national major priorities, high level strategies and practical considerations were heard and integrated in any further developments. Consultations were often used as the most effective tool for implementing this collaborative model. One specific component of the engagement process in this Task T8.1 has been the development of the Specialisation Plan which provides an overarching and comprehensive picture of the strengths of the different participants and may lead to the identification of relevant centres of excellence in specific issues and/or substantiated service providers for the RI.

5.3.6 National Smart Specialisation Plan

Description

Task 8.1 DiSSCo national nodes engagement (Consortium of European Taxonomic Facilities, CETAF (4.5), Royal Belgian Institute of Natural Sciences, RBINS (4), Naturalis Biodiversity Center, Naturalis (3), Finnish Museum of Natural History Luomus (3), Muséum national d'Histoire naturelle, MNHN (3), University of Copenhagen, UCPH (2), University of Tartu, UTARTU (2), University of Florence - Natural History Museum, UniFi (2), Botanischer Garten und Botanisches Museum Berlin, BGBM (1), Natural History Museum, London, NHM (2), Museum für Naturkunde Berlin – Leibniz Institute for Evolution and Biodiversity Science, MfN (1), Institute of Biodiversity and Ecosystem Research – Bulgarian Academy of Sciences, IBER-BAS (2), University of Warsaw, UW (2), Národní muzeum, NM (2), Natural History Museum Vienna, NHMW (2), Agencia Estatal Consejo Superior de Investigaciones Científicas, CSIC (2), Hungarian Natural History Museum, HNHM (2), Plant Science and Biodiversity Centre, Slovak Academy of Sciences, IBSAS (2), Swedish Museum of Natural History, NRM (2), Natural History Museum, Oslo University, NHM-UIO (2), Natural History Museum of Crete, University of Crete, UoC-NHMC (2), National Museum of Natural History Luxembourg, MHNHL (2), Natural History and Science Museum of the University of Porto, UPORTO (2), M01 - M36): National Smart Specialisation and institution-level strategies inform prioritisation objectives in each DiSSCo member country. Gathering the state-of-the-art nationally **provides the basis for the construction of an overall strategic map that is necessary for DiSSCo activities distribution and operation**. Equally important is to **establish a follow up mechanism** to ensure alignment and harmonisation with national RI roadmap processes and relevant foreseen developments (e.g. national contributions/nodes for EOSC-European Open Science Cloud participation in cluster services development, etc.). This means close relations with all national nodes to channel the engagement of governments (whichever the level might be for each country). Out of these activities, a granular thematic specialisation plan ranging from national to institutional levels is produced. For this to be effective, apart from the participating partners' involvement, each national node receives 2 PMs that go to the institution representing the node (to have due legal partnership).

Timeline and Summary Recommendations

Table 33. National Smart Specialisation Plan Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing													
Pri	**Rec**		2023				2024				2025				2026			
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

#	Description																			
127	The Specialisation tool is delivered as a "tested" tool on a sample of institutions from the NH community and Nodes. Although all the NN currently part of DiSSCo have been contacted the current data entered only represent a limited number of data and thus it will require additional efforts to **obtain from all NNs a representative set of data** leading to the analysis of the specialisations present throughout the community of NH institutions.	I	I	I	I	F	F	F												
128	**Ensure at least two institutions per node provide data.** The Specialisation tool gives access to the characteristics and assets of research institutions and nodes that are part of the community of natural history institutions. It is therefore recommended that at least 2 institutions per node provide their data to be able to have a critical mass of data available to search into and enable the expected analysis of key specialisations. The development of the tool is finished, although it can further evolve and be refined. Its implementation "I" phase is ongoing with enlarged data sets provided from the NH institutions, with a minimum of 2 institutions per NN having contributed to the data set. We consider that an additional year would be needed to reach this level. The Finalisation phase will be reached when the analysis will be applicable to a representative set of institutions of at least 2 institutions per Node.	I	I	I	I	F	F	F												
129	**Manage and maintain the Specialisation tool** by a representative of the network and community of natural sciences institutions. The information held in the database represents key assets that should be used diligently. This information will also be enriched by further users of	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O

		the DiSSCo RI that will rely on the FAIR access and GDPR protocols which will ensure a trustworthy repository.															

Discussion and Conclusions

The Specialisation tool is part of the DiSSCo e-services. Link with SYNTHESYS+ ELViS, CETAF Collection Registry, Policy tool (T7.3), Training Strategy WP2.

Basic recommendations can be extracted from the exercise to develop the Specialisation tool that will lead to a series of actions to 1) further enhance the tool and its use; and 2) embed it in a wider scenario, under CETAF and DiSSCo umbrellas.

The future use of the tool could be integrating it in the recurrent activities of the CETAF/ DiSSCO communities with a regular update of the information related to the various categories.

- · The technical maintenance of the platform can be achieved by the Royal Belgian Institute of Natural Sciences (IRSNB) using the collections servers (Plone, Elastic Search) and a specific Google Drive.

- · The RBINS/CETAF team needs to continue to develop the technical aspects of the tool with the help of Franck Theeten (Africa Museum). It was proposed as an e-service of the Belgian federal level to DiSSCo together with a Convention signed with CETAF for complementary work.

The DiSSCo RI will define priorities and may give guidance towards further modifications of the tool in accordance with new standards such as the Latimer Core for collection descriptions or new needs to align it with the other technical components of DiSSCo.

Method

RBINS as T8.1 task leader has delivered at M26 (initially due M18) MS 8.5 "Initial Findings for the Thematic Specialisation Plan". The report was a first step defining the guidelines and data collection process to be further developed, refined and then enriched with the partners own specialisation and potential service provision for the RI.

Deliverable D8.2 Thematic Specialisation Plan (M32+2 postponed to M34 in order to reach as many NN as possible and obtain a valuable set of data before it could be handed to CSO and CETAF), is developed to enable the data gathering on a large set of key areas and services from partners and National Nodes facilitating the user access to the most efficient, robust and reliable data and services existing among DiSSCo community. During the following months March - July 2022, RBINS both continued developing the Specialisation Tool and multiple search function, and its communication towards more National Nodes,

either through bilateral meetings or via NN meetings, to inform them on the process and use of the tool. It is to note that throughout the process, what was initially entitled "Specialisation Plan" has turned into a "Specialisation tool" because it was made clear that the collection of the data enabling the analysis and mapping of the specialisations existing among DiSSCo partner institutions would go through data collection, databasing and data processing.

The data collection and communication towards the NN is to continue beyond the time of the DPP project and so far only a few NN have contributed fully with their data. The tool is delivered with a brief analysis of the data obtained so far as well as a user manual. The Deliverable consists of: i) the tool operational for DiSSCo, ii) the analysis of the expertise and specialisations already gathered, and iii) the guidelines enabling more users to enter their data.

The objective of this tool is to develop a unified information collection system for all the data needed for the specialisation exercise. This centralised data collection should then allow analyses to be carried out at the level of DiSSCo infrastructure, of each national node or each institution. The data should also be available via web service to external tools or users.

A. Topics addressed by the Specialisation tool:
1. Collections
    a. General list illustrating the diversity of collections in the institution
    b. Detailed list of collections and sub-collections with all the data used to feed the Collection Dashboard developed by NHM (London), the CETAF collection registry or the SciColl (Scientific Collections International) registry hosted by GBIF (Global Biodiversity Information Facility).
2. Research domains
3. The expertise to be found at two levels:
    a. Institutional
    b. Individual
4. Infrastructures
    a. Research infrastructure
    b. Scientific instruments
5. Exhibitions
    a. Permanent galleries
    b. Temporary exhibitions (proposed as product to external clients)
6. Training courses delivered
B. Standardisation of the tool
    · ROR (Research Organization Registry) identifiers are used for institutions

- ORCID (Open Researcher and Contributor ID) identifiers are used for individuals
- JASC 3 (Joint Academic Coding System) categories are used for research topics

The tool also uses controlled vocabulary lists and/or pre-encoded options whenever possible

C. Technical approach

    1. Data collection

The objective behind the technology that has been developed was to enable:

    a. the entry of data into specific forms created per national node with associated rights to the person responsible for each node.
    b. the entry of data from existing systems, e.g. via XLS templates
    c. the system to evolve with new fields or new options without having to completely revise the tool at each update
    d. data aggregation at the level of each national node to the level of the infrastructure producing a global analysis
    e. a simple data storage in spreadsheet format
    f. the update of data based on previously encoded information
    g. GDPR compliance

According to these criteria, it was decided to use Google Forms / Google Sheets placed in a dedicated drive space.

Data can be entered directly by the NN or institution either via forms or via XLS files made available respecting the same structure. However, the use of forms is recommended as it is much better at handling multiple choices and the protection of personal data.

The data collected is stored directly in Google sheets which can be consulted by each national node.

A total of 18 forms (9 data forms and 9 forms for importing XLS files) were created for each national node, i.e. 414 forms for the whole infrastructure.

A series of scripts were developed to maintain these forms:

- updating the institutions within each national node
- modify the options for each single choice or multiple value field
- modify field headings

    2. Data aggregation

Specific scripts have been developed to aggregate the data collected by each national node:

· Scripts for the Google sheets fed by the forms

· Scripts for data manually encoded in XLS templates. These data must be manually checked before merging with the data produced by the forms.

3. Data analysis

A copy of the aggregated data of each topic is created to perform the analyses. For each topic, a specific google sheet gathers the aggregated data, a copy of the cleaned data and different sheets with pre-built graphs that allow to visualise the total number of answers, a distribution map of the answers and the main data related to the topic.

Each graph has its own "url" which allows it to be included in any website that accepts Iframes or embedded objects. This guarantees continuity of data use even if the website technology changes and even if the websites displaying the graphs (e.g. CETAF, DiSSCo, National Node site) use different technologies (Google sites, Wordpress, Drupal, Plone etc,...).

4. Saving data

All data is stored as google sheets which can be saved by the national node as XLS or CSV enabling its use in a third party application or for long term storage.

5. Indexing in ElasticSearch indexes

The data is imported into different ElasticSearch indexes to allow quick access via web services.

6. Updating of data

Each completed form has a corresponding google link which is sent by email to the email address indicated in the form. This allows the data to be updated and corrected before their final validation and use.

A complete backup is made before each form is sent to update the data. A pre-filled form with the data already encoded is sent to the email address provided for periodic updating of the data used as a basis for the analysis.

D. Workflow of the specialisation tool
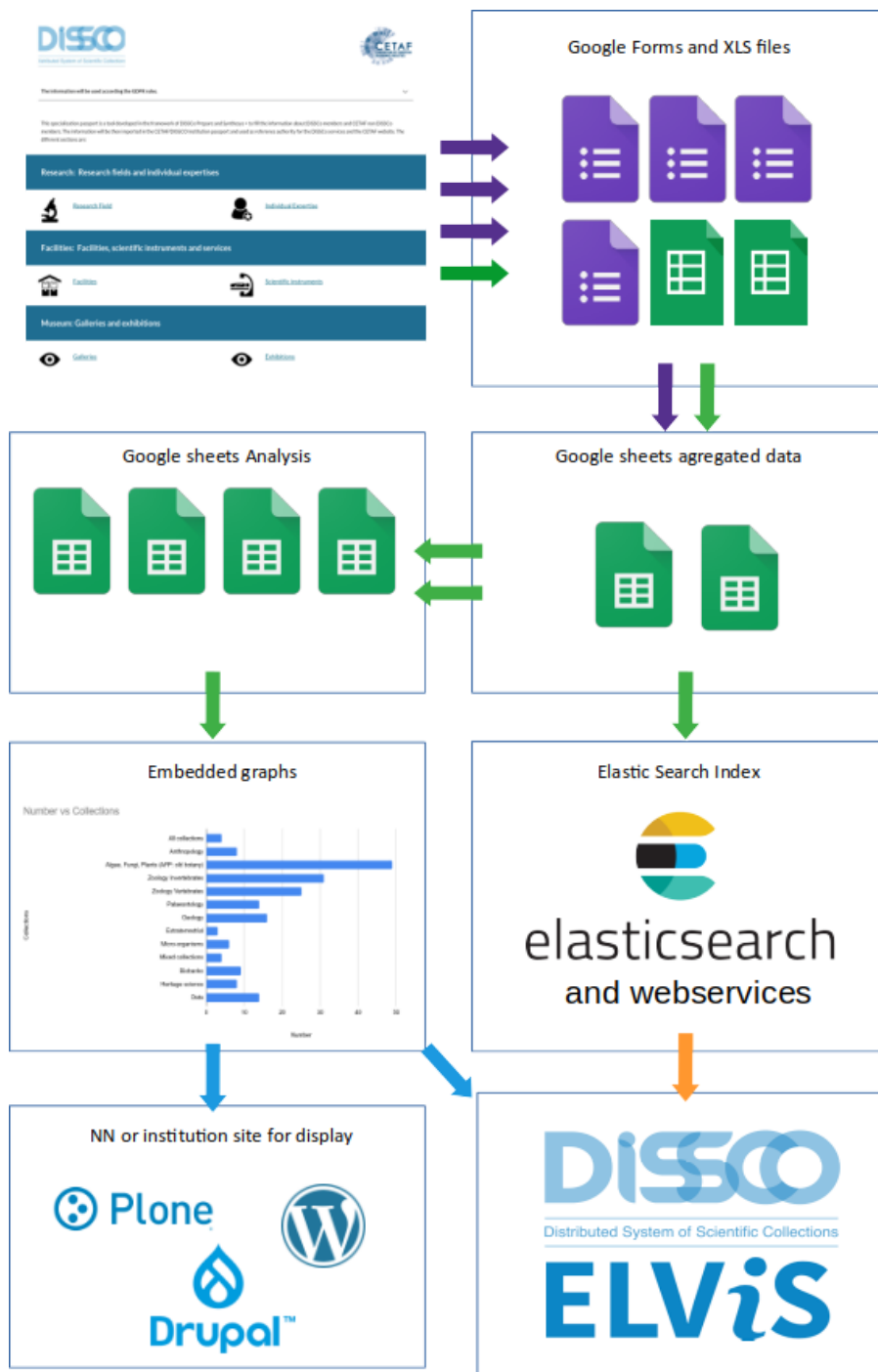
E. Structure of each NN folder in the Google drive

5.3.7 Communication and Engagement

External Communication Strategy

The task aimed to set out a concise communication plan, employing a four-layered approach for targeting relevant audiences, and leverage other DiSSCo-related projects' (i.e ICEDIG, SYNTHESYS+, and MOBILISE) targeted audiences: (i) direct contributors in the broader network of collection holders; (ii) complementary stakeholders and initiatives (including international initiatives/RIs such as iDigBio, CRIA, ALA, GBIF, in the environmental and digital domains); (iii) related third parties (users platforms, e.g. EOSC, European Open Science Cloud); and (iv) societal bodies (including governmental representatives and funding agencies at regional, national, European, and international scale, European Strategy Forum on Research Infrastructures (ESFRI) delegates, media and the general public).

Based on the Communication Strategy, several external communication tools have been developed and operated. These include a logo and graphic charter (developed under WP9 - Project Management), the project website (which is hosted as part of the DiSSCo website – dissco.eu) as well as social media channels. All project partners have contributed to feeding any newsworthy content into these channels and many of them have actively engaged with their respective communication and press teams.

In terms of communication, DiSSCo has evolved from a static picture to a more dynamic environment thanks to the different tools created and exploited throughout the lifetime of the preparatory phase to highlight primarily a) the shared understanding of the operation of DiSSCo and the way the DiSSCo community can contribute to its construction; b) the capacities and benefits of DiSSCo for different communities of users; c) the added value of the DiSSCo planned services for tackling biodiversity crisis and achieving Sustainable Development Goals; and d) the socioeconomic impact, the scientific dimension, and innovation push factors that DiSSCo encompasses once operational.

Timeline and Summary Recommendations

Table 34. Communication and Engagement Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	130	**Revise and adapt the Communication Strategy according to the next phases of DiSSCo.** Communication work is a non-stop task that necessarily will evolve as DiSSCo develops, and needs to be customised		P	P	I	F											

		according to the RI's evolution. Therefore, actions in communication need to be maintained and furthermore, increased during the construction phase towards the operation launch of DiSSCo.																	
131		**Define and implement an evolving schema for communication material** so that users and interested agents can easily understand the status-quo of DiSSCo, the progress made, and the outstanding stages to address;	P	P	I	I	I	I	F									
132		**During Transition further develop and put into practice, the figure of DiSSCo Ambassador,** outlined during the preparatory phase to address strategically the constitution of the DiSSCo ERIC and later on, during the construction phase, the expansion of its membership;	P	P	I	I	I	I	I	I	I	I	F				
133		**Encourage increased alignment with sister initiatives in Europe and globally to continue the ongoing processes of linking RIs acting in the field,** through either joint projects such as BiCIKL, or collaborative strategic partnerships			P	I	I	I	I	I	I	I	I	I	I	I	I	F
134		**Maintain a person fully dedicated to DiSSCo communication tasks** to keep the targeted audiences duly updated and well informed, and to embed these activities also into the development of the DiSSCo ERIC.	P	P	I	I	I	I	F									

Discussion and Conclusions

With the contributions from partners, and under the direct supervision of the CSO, the communication team has developed strong and focused material that constitutes important tools for outreaching and disseminating the DiSSCo concept. Some of those intend to address particular topics as is the case of **DiSSCo Binnacle** that offers a very detailed and comprehensive picture of the different services that DiSSCo aims to provide. Others target a broader audience to e.g. illustrate the benefits of DiSSCo to the scientific community (DiSSCo key messages), provide specific supporting documentation to bodies (Package for the FF), and present the content and usability of the different DiSSCo community services

(specifically e-services). Finally, other types of supporting material have also been created to be used for presentations, events and ultimately at any dissemination scenario where DiSSCo may participate (i.e outline, brochure, leaflet, presentations). All of these tools have been compiled at the DiSSCo website, under "Resources[158]". Dedicated social media channels have been key in punctually raising awareness and disseminating the DiSSCo activity and progress.

Method

The communications activities have been strongly anchored and supported by the CSO thus enabling adequate alignment with actions undertaken by other DiSSCo-related initiatives and the CETAF community at large. A dedicated Communication Officer has overseen the implementation of the Communication strategy in close collaboration with the CETAF Communication Officer in weekly meetings. Such a collaborative mechanism has facilitated having harmonised calendars with alternate newsletters (DiSSCo Highlight and CETAF Pulse) and combined dissemination campaigns, presence in conferences and coordinated participation in events. Several communication materials have been produced and disseminated to respond to the detected needs of the community as well as those arising from community meetings, either at the Business stream level, or as a National Node request, or equally in support of the different DiSSCo governance bodies (iGA or FF) or advisory boards (SAB and TAB). Moreover, as a specific demand from the iGA, specific high-level working teams have been established to address specific topics as has been the case for the revision of the vision, mission and values of DiSSCo, and an analysis run to identify interlinkages among different biodiversity-related RIs ("Contact zones exercise"). The first team was led by Edwin van Huis, the iGA Chair, and the second by Aino Juslén, iGA Vice-Chair.

5.3.8 Strategic Partners Engagement

DiSSCo stakeholder engagement

Description   *DPP D8.3*

Built on the work performed by DiSSCo-related initiatives, Task 8.3 strengthens the linkages and builds strategic partnership frameworks with the relevant external stakeholders and communities.

The goal of these engagement activities is to identify synergies, coordinate actions, and develop joint services. Alongside target audiences identified in Task 8.1, DiSSCo National Nodes Engagement, this activity specifically addresses stakeholders in three axes: a) legal:

from private (IT partners and other industrial actors) to public (e.g. academia, citizen science); b) field of reference: domain-specific to related entities (e.g. Biodiversity Heritage Library-BHL, Open Access publishers as well as RDA-Research Data Alliance, and ECSA-European Citizen Science Association, and c) geographical spread: from European to global international coverage.

This task looks at procedures for DiSSCo external collaboration and tendering, both collectively and individually. DiSSCo facilities partnered and contracted with third parties to obtain, e.g. computing capacity and to deliver collections management, digitisation, training or publishing, etc. This task sets out procedures and procurement frameworks, to ensure consistent practice and maximise successful outcomes across the consortium. It also considers the wide variety of potential partners and contractors, from businesses of all sizes, to publicly or philanthropically funded organisations, social enterprises, and research organisations.

The output of this task is a framework for alignment and cooperation and a set of best practices for partnerships to act with partners as described above, but with other relevant European RIs and initiatives included (towards EOSC, but also ERA-European Research Area, ESFRI, Europeana, EuroGEOSS-Global Earth Observation System of Systems, in addition to eLTER-Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research, LifeWatch, E-RIHS-European Research Infrastructure for Heritage Science through the clustering ENVRI FAIR project, etc.).

Timeline and Summary Recommendations

Table 35. Strategic Partners Engagement Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	135	**Procurement Strategy:** The **MS8.4 milestone** recommends that DiSSCo develops a procurement strategy in order to align long term priorities and objectives with procurement processes. **This strategy should help DiSSCo to develop partnerships, scale up processes, as well as mitigate risks and improve cost efficiency.** This work is continued by T4.4. **Stakeholder Analysis: MS8.6** details																	

		DiSSCo stakeholders, including GBIF, Catalogue of Life, GeoCASe, TDWG and International Barcode of Life. **This report informs stakeholder engagement plans and prioritisation.**															

Discussion and Conclusions

Procurement Strategy: The MS8.4 milestone recommends that DiSSCo develops a procurement strategy in order to align long term priorities and objectives with procurement processes. This strategy should help DiSSCo to develop partnerships, scale up processes, as well as mitigate risks and improve cost efficiency. This work is continued by T4.4.

Stakeholder Analysis: MS8.6 details DiSSCo stakeholders, including GBIF, Catalogue of Life, GeoCASe, TDWG and International Barcode of Life. This report informs stakeholder engagement plans and prioritisation.

Method

Procurement Strategy (MS8.4)[159]: This milestone consists of a policy briefing on key areas of procurement strategy, and links closely with T4.4 Pre-commercial procurement.

Stakeholder Analysis (MS8.6)[160]: Stakeholder analysis workshops took place with task partners from T8.3 and WP4 Business Framework. These workshops identified DiSSCo stakeholders within the bio/geodiversity, environmental and life sciences related research infrastructures and organisations. The stakeholder analysis was supported by the infrastructure contact zones survey, which is an analytical framework developed by a DiSSCo General Assembly task force.

Stakeholder Engagement: Stakeholder engagement plans are included in the final deliverable for this task.

5.3.9 Advocacy and Engagement

National and European governments dissemination, outreach and advocacy

Description   *DPP D8.1*

The decision of the DiSSCo interim General Assembly (iGA3) in June 2021 to adopt the ERIC model as the future DiSSCo legal entity implied a new governance stipulated by the provisions of the new DiSSCo legal entity Statutes. The legal entity will depend on national political and financial commitments while still operating through institutions organised under national nodes.

Such a transition from facilities-based governance to country formal engagements demanded a structured process to guarantee an active engagement of governmental representatives at an early stage of the preparation of the RI.

To do that, DiSSCo established the National Nodes playground where the National Nodes play a critical and multi-purpose role in terms of external engagement: 1) keeping digitisation plans envisaged at a national level aligned with DiSSco strategy and architecture; 2) facilitating smooth channelling of necessary information from DiSSCo facilities to governmental representatives; and 3) promoting the value proposition of DiSSCo to articulate the right messages from the Research Infrastructure and customise them to accommodate to specificities of the country where the National Nodes is established. Moreover, the efforts in obtaining critical commitment from governments and funding agencies were deployed consistent with a defined advocacy strategy whose primary objective was to mobilise DiSSCo node representatives in communication and advocacy campaigns that would facilitate setting-up the Funders Forum (FF). The role of the FF is to establish an open and productive dialogue with representatives of the future member state funders of the infrastructure.

The FF comprises representatives from eleven countries (The Netherlands, Italy, France, United Kingdom, Belgium, Greece, Estonia, Bulgaria, Slovakia, Denmark and Portugal). The operation of the FF1:

1. provides valuable guidance to the interim bodies regarding preparatory phase activities of the infrastructure;

2. improves the understanding of DiSSCo by the future funders generally in the infrastructure-related and data-intensive research pan-European landscape and also locally in the national context, and

3. establishes trusted discourse mechanisms across interim (Preparatory Phase) and future (Implementation and Operation phase) DiSSCo management and governance bodies.

More specifically, the operation of the FF, during the Preparatory Phase, has allowed for a continuous consultation on relevant operational, organisational and engagement matters. Inputs from the FF guided the delivery of key outcomes for the RI during the preparatory phase and informed relevant decisions for the development of DiSSCo. This is the case of the decisions on the DiSSCo ERIC governance model, DiSSCo business framework, DiSSCo Statutes or the DiSSCo ERIC roadmap.

The advocacy strategy achieved was also another successful objective, establishing fluent communication of DiSSCo nodes with national funders that positively impacted further development of existing Nodes (i.e. Spain), inclusion in national roadmaps (i.e. Bulgaria, Estonia) or the setting up of agreements around the facilities at national level (i.e. UK).

Advocacy continues being a constant during the implementation phase.

Timeline and Summary Recommendations

Table 36. Advocacy and Engagement Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing																
Typ	No	Recommendation	2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	136	**Revise and adapt the Advocacy and Engagement Strategy** according to the next phases of DiSSCo.	P	P	I	I	I	I	I	I	I	I	I	I				

Discussion and Conclusions

Advocacy continues being a constant during the implementation phase.

A methodology based on design thinking that has proved to be an excellent mechanism to develop DiSSCo can be used to improve the strategy, including new activities and new stakeholders such as strategic partners.

Method

It was important to equip the DiSSCo Research Infrastructure with mechanisms to avoid misalignments with the national scientific priorities and guarantee long-term sustainability in the scientific, technical, governance and financial dimensions.

With this objective in mind, DiSSCo was in the early stages of working towards sound commitments at a national level across 23 countries already engaged with the research infrastructure by tooling national nodes to establish sound long-term communication channels and to convey the right messages to the target audience.

Common narratives were presented and discussed at the NNs meetings regularly and other communication tools. Bilateral meetings and surveys were carried out to collect relevant information to tailor-made national discussions. The information is part of a database that is updated regularly and used by the DiSSCo CSO for multiple purposes.

06 Finances

6.1 Overview
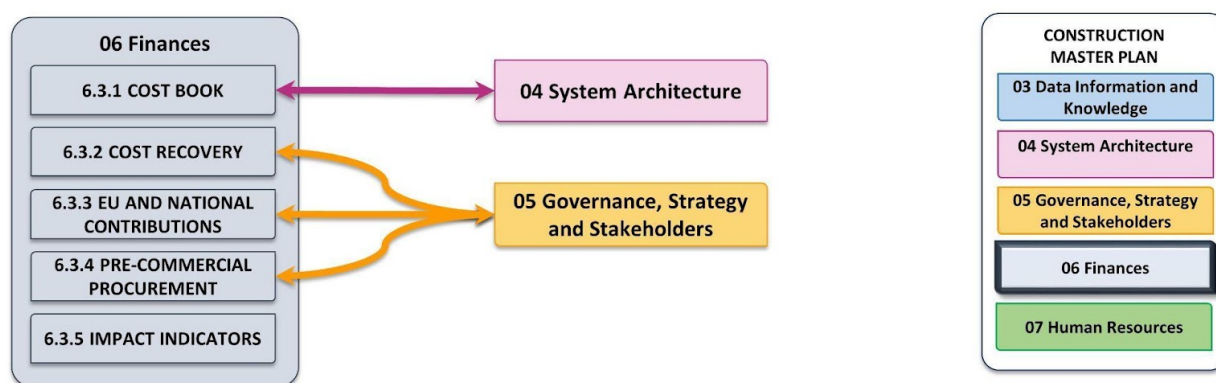
Figure 8. Synoptic graph showing the dependencies (arrows) between tasks in this chapter with other tasks or chapters (systems) in DiSSCo Construct.

Understanding the financial parameters of a highly distributed international organisation is challenging, as there are numerous factors that constantly evolve and can affect the financial health of the organisation. The challenge becomes even greater when we take into account the mix of in-cash and in-kind contributions (income) from multiple sources that DiSSCo infrastructure relies upon. During the DiSSCo Prepare Project we focused on documenting the financial instruments and parameters DiSSCo must rely upon during the DiSSCo Construction and Operational Phases. The work focused on delivering a robust cost book with a high level of granularity that covers the key cost categories of DiSSCo. We linked the contribution model of the member states to that cost book and investigated innovative instruments that can provide additional financial support or improve financial performance, such as pre-commercial procurement.

The coming steps underlie the preparation of the European Research Infrastructure Consortium (ERIC) application process. As part of the documentation to submit in Step 1, the Host Country and Founding Members negotiate the future membership fee and related matters.

To prepare negotiations, the Funders Forum has to decide on the final version of the draft to be handed over to the ERIC members. The first discussions took place in the Funders Forum 5th meeting (December 2022) and continued during the Funders Forum 6th meeting (early Spring 2023).

The work has also provided a clear insight into a set of socioeconomic indicators for the assessment of the research infrastructure and continuous improvement.

Further actions are not foreseen during the first year of the transition phase on additional financial support (i.e. Pre-commercial Procurement-PCP). Those and other matters are the responsibility of DiSSCo-ERIC to decide upon.

6.2 Timeline and Summary Recommendations

Table 37. Finances Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

				"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
	Pri	Rec		2023				2024				2025				2026			
Task Name	Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
6.3.1 Cost Book	Tr	137	**(H) Design a training for cost estimation.** Should be designed before construction phase and implemented in 2024 → NN representatives fully trained by 2026.	P	P	I	I	I	I	I	F	F							
6.3.1 Cost Book	H	138	**Develop a web platform to ease cost assessment.** Can be developed during the construction phase.	P	P	I	I	I	I	I	F	F							
6.3.2 Cost Recovery	H	140	**Develop a Financial Sustainability Plan** including income diversification.	P	P	P	P	I	I	I	I	I	I	F	F	F	F	O	O
6.3.2 Cost Recovery	H	141	**Do not charge academic researchers, nor users from member and observer states for commissioned services.**																

			Most ERICs charge users from other states and users from industry to access their services.																
6.3.3 EU and National Contributions	H	145	**Produce an annual report: linking KPIs** (key performance indicators) **and annual budget**	P	P	P	P	I	I	I	I	I	I	I	I	I	I	I	I
6.3.3 EU and National Contributions	H	146	**Keep an eye on EU funding calls / opportunities**	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
6.3.3 EU and National Contributions	H	147	**Engage new Member States** to increase annual national contributions and DiSSCo fixed funding	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
6.3.4 Pre-Commercial Procurement (PCP)	H	149	**Update the DiSSCo Construct development needs (H, prep needed)**: By listing the development needs according to their industry TRL' s and their fit with the 4 strategic pillars defined during DiSSCo Prepare, we can create a list of eligible activities and their priorities, that can then serve as pipeline for PCP and PPI opportunities. a first assessment and selection has been completed, this should be repeated at set intervals (e.g. quarterly), when an intermediate perspective is compiled centrally.								P	P	I	I	I	F	F	F	F
6.3.4 Pre-Commercial al	H	150	**Establish DiSSCo as a legal entity (prep needed)**: In order to leverage the benefits of PCP and PPI during the Construct Phase of DiSSCo, we need a legal entity or one of							P	I	F	F						

Area		#	Description																	
Procurement (PCP))			the consortium partners to assume the procurement role during the construct phase. At which time, also the procurement structure in DiSSCo Hub should be locked.																	
6.3.4 Pre-Commercial Procurement (PCP)	H	151	**Lock on DiSSCo procurement structure (H, prep needed)**: the procurement function can be centralised, decentralised, hybrid,.. Each setup has its own merits and drawbacks, nevertheless, a timely choice would benefit the future efforts in DiSSCo.								P	I	F	F						
6.3.5 Impact Indicators	H	154	**Identify detailed data requirements** - Liaise with technical development and partners to provide detailed data requirements on monitoring services, tools, activities to ensure implementation of data gathering and data flow			I	I	I	I	I	I									
6.3.5 Impact Indicators	H	155	**Establish a pilot study of indicator quantification** to establish background status of SEI indicators before operation					I	I	I	I									
6.3.1 Cost Book	M	139	**Organise a meeting with national nodes to share / review the area concepts.** Can be developed during the construction phase.	P	P	I	I	I	I	I	F	F								
6.3.2 Cost Recovery	M	142	**Diversify sources of income for sustainability of the ERIC** by: (a) creating interactions and partnerships between the ERIC and different stakeholders, both in the public and private sectors, at several levels (International, European, national and regional), and (b) diversifying the services																	

Section			Recommendation																			
			provided by DiSSCo-ERIC and apply a charge to access them.																			
6.3.2 Cost Recovery	M	143	**Develop a Central Business Development team to further develop a cost recovery framework** including trialing some of the proposed business models, and managing a portfolio of income diversification projects and services.																			
6.3.2 Cost Recovery	M	144	**DiSSCo ERIC should not be constrained in charging for services in its statutes** when specifying a maximum level of income from economic activities.																			
6.3.3 EU and National Contributions	M	148	**Provide** <u>training</u> **regarding submitting funding applications and consortium building**	P	P	P	P	I	I	I	I	I	I	I	I							
6.3.4 Pre-Commercial Procurement (PCP)	M	152	**Issue a Prior Information Notice (PIN) (M, provided the pre-commercial procurement high priority recommendations have been completed)**: with a refined scope of the proposed development, performing a gap analysis to track existing solutions and their TRL levels to define the preferred procurement strategy (PCP, PPI, standard procurement),									P	P	I	I	F	F				
6.3.4 Pre-Commercial al	M	153	**Work out the PCP and PPI contract notices (M, provided the pre-commercial procurement high priority recommendations have been completed)**: Engage in the preparatory work for organising PCP and PPI calls, which									P	P	I	I	F	F				

Procurement (PCP)			includes a clear scope definition, preliminary market consultation leading up to a publication of the contract notice, which details the problem to be solved, its functional requirements and use cases.																
6.3.5 Impact Indicators	M	156	**Define roles within the DISSCo organisation about data gathering, analysis and reporting of SEI indicators**																
6.3.5 Impact Indicators	M	157	**Start the design and implementation of the information management system for data gathering, analysis, reporting and archiving**. It might require new developments to implement automatisms;																

6.3 Task Summaries

6.3.1 Cost Book

Cost Book for DiSSCo

Description   *DPP D4.1*

OBJECTIVES

The DiSSCo Cost Book provides the financial elements to serve both internal and external purposes. DiSSCo operations must incorporate knowledge about the operating costs of maintaining the Information Technology (IT) infrastructure and the associated services. It must also be able to evaluate the cost of the service portfolio it provides in order to charge the appropriate costs to its users.

The core service of DiSSCo is the access to information within two main areas: access to existing digitised information and digitisation-on-demand.

The cost of access to existing digitised information depends on choices made for storing data, which controls data mobilisation speed. DiSSCo governance should be able to prioritise sets of collection data requiring fast access based on users' needs and relevant impact assessments, as identified in DPP Work Package 1.

The cost of digitisation-on-demand depends heavily on resources for access to physical specimens. DiSSCo operations should incorporate knowledge about the cost of accessing, hence digitisation, specimens for each member, and possibly for each scientific collection department. At the DiSSCo level, it can propose to the users a programme of digitisation that creates the set of data it needs.

The parameters which are developed also serve as performance indicators for DiSSCo business operations. The digitisation-on-demand process must be as thorough and fast as possible; therefore, the speed of access is a key element. Based on access speed, DiSSCo business operations should be able to establish priorities with respect to current trends of users' interests. Establishing the cost book faces several challenges due to the diversity in DiSSCo members; different sizes, organisations, methods of accounting, making it difficult to identify the cost of DiSSCo activity which is generally a fraction of the DiSSCo members activities. Work Package 4 aims at delivering the DiSSCo cost book in a way that is as much as possible independent of institution size and countries and that covers all aspects of

funding, including the design of a sustainable business model. For this purpose, WP4 is divided into four tasks.

DiSSCo Prepare T4.1 is dedicated to DiSSCo research infrastructure Cost Book calculation. It aims at identifying cost categories and to calculate the figures associated. Such a task covers the construction, the operation and the decommissioning phases of the DiSSCo Research infrastructure. In practical terms, costs are distributed among different categories: (i) the Central Hub costs: coordination, management, administration, IT systems, development of services; (ii) decentralised costs: cost to operate DiSSCo associated services (digitisation, physical access, consulting services, training, loans). Such a division is due to different resource persons related to these tasks.

Subtask 4.1.1 – Cost indicators and methodology - MNHN

The first task of this subtask is to design a methodology that allows institutions participating in DiSSCo to better estimate their tariff for the services they could provide with DiSSCo RI. In order to develop the method, the Muséum National d'Histoire Naturelle (MNHN) first delimited the perimeter of the RI and defined 14 cost areas. These areas correspond to 3 types of category of costs: 1) cost of the services provided by the RI; 2) support cost to preserve natural history collections and data associated; 3) cost to participate and coordinate national nodes / EU projects / data management.

The second task of this subtask is to work with DPP work package leaders in order to better understand what is the output of their WPs. This allows WP4 to design a clear perimeter for the RI and both estimate the costs of the hub and the costs of the nodes. The output expected is a clear framework of the activity of the DiSSCo RI and an estimation of their costs.

Subtask 4.1.2 – Costs of DiSSCo services - NHM

T4.1.2 outlines the options to support the following planned and in development e-services: Authentication and Authorization Infrastructure (AAI); Collection Digitisation Dashboard (CDD); Digital Specimen Repository (DSR); European Loans and Visits System (ELViS); Helpdesk; Knowledgebase (KB); Specimen Data Refinery (SDR); Unified Curation and Annotation System (UCAS).

Services are prioritised based on risk a)nd Technical Readiness Level (TRL). T4.1.2 obtains rough cost estimates for Development phases (through TRL stages), Technical maintenance (including hosting and other ongoing costs) and Helpdesk support and/or training. T4.1.2 considers different financial sustainability models and when they are best used.

T4.1.2 also considers more detailed costing for digitisation services, building upon the work from the Innovation and Consolidation for large scale DIGitisation of natural heritage (ICEDIG) project[1]. This is based upon the additional needs from the T4.1.1 Cost Book. This may include understanding data mobilisation, data processing activities and data access and exploitation costs, where this has not already been included in the Cost Book.

Subtask 4.1.3 – Detailed costs of IT infrastructure – MNHN and SGN

The perimeter of this subtask depends on WP6 and the DiSSCo Coordination and Support Office (CSO) technical team. The way DiSSCo technical architecture is designed impacts its construction and operation costs. The approach here is to better understand who is responsible to manage and preserve DiSSCo data, the design of the distributed architecture, what are the digital services of the RI, the division of responsibilities and the interfaces between central and local components and what is the investment needed to implement and maintain them? A first step is to clarify this perimeter with DPP partners. A second step is to identify the costs associated and if possible, provide figures

Timeline and Summary Recommendations

Table 38. Cost Book Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

| | | | "P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing | | | | | | | | | | | | | | |
| Pri | Rec | | 2023 | | | | 2024 | | | | 2025 | | | | 2026 | | | |
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	137	**(H) Design a** <u>**training**</u> **for cost estimation.** Should be designed before construction phase and implemented in 2024 → NN representatives fully trained by 2026.	P	P	I	I	I	I	I	F	F							
H	138	**Develop a web platform to ease cost assessment.** Can be developed during the construction phase.	P	P	I	I	I	I	I	F	F							
M	139	**Organise a meeting with national nodes to share / review the area concepts.** Can be developed during the construction phase.	P	P	I	I	I	I	I	F	F							

Discussion and Conclusions

Task 4.1 is about calculating DiSSCo core and associated costs. It is understood that it is not the area of expertise of DiSSCo stakeholders and, at the same time, it is a fundamental

brick to speak with funders. If DiSSCo is able to present an accurate budget, it demonstrates robustness to its funders.

In terms of dependencies, WP4 relies on ICEDIG previous tasks, notably:

· <u>Deliverable 4.5: Cost analysis of transcription methods</u>[161];

· <u>Deliverable 8.1: Conceptual design blueprint for the DiSSCo digitisation infrastructure</u>[162];

· <u>Deliverable 8.2: Cost Book of the digitisation infrastructure</u>[163].

From a larger perspective, T4.1 relies on the DiSSCo masterplan and what the project aims to concretely implement in the following years. The outputs of WP9 represent an actionable plan for the RI and at the same time, concrete costs. As due dates for deliverables do not match, T4.1 relies on WP and Task leaders of DiSSCo Prepare to estimate what are the concrete costs to operate the RI.

Finally, T4.1 relies on the willingness of national nodes to participate in the costing methodology: assess the costs of their institutions and share the methodology with their members.

Method

The DiSSCo cost book is based on two main blocks:

1. Calculation of the costs to build and operate the research infrastructure: interviews with WP and task leaders to forecast the costs DiSSCo ERIC represent.
2. Methodology shared with natural history institutions to help them calculate the costs associated with the services they could provide as part of DiSSCo: consulting services; mass digitisation; small set digitisation; training; loans of collections; physical access to collections.

6.3.2 Cost Recovery

Cost model for charging services

Description *DPP D4.2*

This task is built on the user stories and case studies compiled in WP1 (<u>Tasks 1.1 and 1.2, Life Science and Earth Science Use Cases</u>) and the user needs identified in Task 8.3 <u>Strategic Partners Engagement</u>. In conjunction with the service needs provisionally identified during the proposal stage of DiSSCo, Task 4.2 identifies and fully costs the range

of services to be provided by the DiSSCo RI. These services support both the institutional data contributors as well as the external user community (e.g. researchers and citizen scientists).

Task 4.2 develops a technical framework to measure these costs and collates these data from selected DiSSCo facilities. Services that need to be costed, broadly fall into the following categories:

· Data mobilisation (i.e. physical digitisation across a range of collection types and service providers);

· Data processing activities (i.e. cleaning and preparation to make the data fit for use);

· Data access and exploitation (i.e. the actions required to supply data to users and track its impact).

In addition to these data services, DiSSCo incurs running costs fundamental to the provision of common services core to the operation of the DiSSCo RI. Examples include:

· Running the DiSSCo Helpdesk (i.e. operating 1st - 3rd line support across for the DiSSCo facilities);

· Training provision (i.e. providing access to training services across all partners);

· Common communication services (i.e. provision of IT services necessary for facilitating the efficient operation of the DiSSCo community);

· Common human resource services (e.g. the costs of secondment procedures).

Task 4.2 develops detailed costs for the provisioning of these services at reference institutions that represent the diversity, geographic spread and size of those present across the consortium. As part of this task, we seek agreement on the platform to store this data, as well as the security and privacy arrangements that govern access to this information. These are developed in conjunction with WP 9 that covers the ethical aspects of DiSSCo development and the technical subcontractor aiding the development of the DiSSCo Cost Book (with Task 4.1).

A technical subcontracted partner helps to identify users and market niches. It then explores with potential users their needs and identifies the key elements to propose or to implement into services in order to fulfil these needs and ensure that the offer matches demand. The technical partner runs some pilot/experimental projects to assess the cost of services. These pilots lead to a full business model with project marketing including the range of offers and prices, and communication strategies. The technical partner liaises with WP1.

Timeline and Summary Recommendations

Table 39. Cost Recovery Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	140	**Develop a Financial Sustainability Plan** including income diversification.	P	P	P	P	I	I	I	I	I	I	F	F	F	F	O	O
H	141	**Do not charge academic researchers, nor users from member and observer states for commissioned services**. Most ERICs charge users from other states and users from industry to access their services.																
M	142	**Diversify sources of income for sustainability of the ERIC** by: (a) creating interactions and partnerships between the ERIC and different stakeholders, both in the public and private sectors, at several levels (International, European, national and regional), and (b) diversifying the services provided by DiSSCo-ERIC and apply a charge to access them.																
M	143	**Develop a Central Business Development team to further develop a cost recovery framework** including trialing some of the proposed business models, and managing a portfolio of income diversification projects and services.																
M	144	**DiSSCo ERIC should not be constrained in charging for services in its statutes** when specifying a maximum level of income from economic activities.																

Discussion and Conclusions

Task 4.2 resulted in two separate reports:

1) "Future opportunities for DiSSCo's financial sustainability"

2) "Report on ERICs financial activities and commissioned services"

The first report contains a summary of the second and provides a preliminary identification of users and of market niches for potential DiSSCo services, to provide recommendations for future scoping and capacity building.

It summarises potential income diversification and the legal constraints on charging for services for DiSSCo-ERIC. Our compilation of the official statutes and annual activity reports of several ERICs, and the results of a consultation on the economic activities of eight ERICs, have shown the following:

· ERICs are encouraged to diversify their sources of incomes: it is fundamental for the sustainability of DiSSCo-ERIC to secure sources of income that fund its activities in the long term. Ideally income streams not only originate from European funding, and are as diverse as possible.

· Two efficient ways to diversify the sources of income of ERICs are:

1. to create interactions and partnerships between the ERIC and different stakeholders, both in the public and private sectors, at several levels (International, European, national and regional);

2. to diversify the services provided by DiSSCo-ERIC and apply a charge to access them.

Both measures should increase the visibility of DiSSCo-ERIC and favour securing diverse income streams.

As long as the economic activities remain limited and do not jeopardise the achievement of the main tasks of DiSSCo-ERIC, there is no further constraint on charging for services in the legal texts framing ERICs. Other ERICs have not mentioned a maximum level of income from economic activities in their statutes, and it is suggested that DISSCo-ERIC aligns with them. Two ERICs have declared aiming for 5% of income from commissioned services. Other consulted ERICs do not have an aim in terms of income from commissioned services but some may develop one in the future. Therefore 5% of income from commissioned services constitutes a basis for a model for DiSSCo-ERIC.

We recommend caution when deciding which users are charged for commissioned services. Our general recommendation is to not charge academic researchers, nor users from member and observer states (considering that the latter do pay an annual fee). Most ERICs charge users from other states and users from industry to access their services.

The sustainability of DiSSCo-ERIC (encompassing income diversification) could be organised within a 'sustainability plan' that could include the measures presented above and could be developed during the construction phase of the DiSSCo-ERIC.

Finally, we propose a Central Business Development team that would continue this work, trial some of the proposed business models, and manage a portfolio of income diversification projects and services.

Method

The majority of Task 4.2 was done by a subcontracted partner (AcrossLimits) with research on ERICs and their statutes undertaken by Gaël Lymer of Royal Belgian Institute of Natural Sciences (RBINS). The market opportunity for DiSSCo Services was explored through a series of surveys, workshops and interviews. The task reviewed legal limitations to selling DiSSCo e-services, and providing service charging options, and a pricing model.

6.3.3 EU and National Contributions

National contributions to the DiSSCo RI

Description   *DPP D4.3*

Task 4.3 aims to develop a sustainable national funding model for the research infrastructure. The working hypothesis is that DiSSCo is an ERIC, which means that it relies on Member states / national governments who are the designated members at the DiSSCo General Assembly. The approach is to work in close collaboration with WP7 and WP8 in the areas of governance and strategy, as the engagement of partners is a core issue for funding.

Task 4.3 has proposed 3 models for government funding and 2 models on how funding could circulate inside the RI. By the end of DPP, the task should stabilise and finalise these models, offering recommendations on the most suitable model, with assistance from future funders and/or subcontractors.

T4.3 also heavily depends on the negotiations of the ERIC statutes. There is a section dedicated to financial provisions and "principles for contributions of Members / observers / intergovernmental organisations". This section both depends on and impacts T4.3.

According to this work, final steps for T4.3 include a continued study of the funding mechanisms in the different DiSSCo countries in order to better construct a globally-endorsed model. It looks with greater depth into the European funding landscape in order to list which EU funding DiSSCo could benefit from, and provides estimated figures for national contributions in order to have an idea of the potential initial budget of the ERIC. Task 4.3 aims to continue to strengthen the communication with DiSSCo national nodes in order to establish a solid communication link with potential national funders. Particular

consideration is given to a contribution involving centres of excellence, and the implications this could have.

Timeline and Summary Recommendations

Table 40. EU and National Contributions Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing																
Pri	Rec		2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	145	**Produce an annual report: linking KPIs** (key performance indicators) **and annual budget**	P	P	P	P	I	I	I	I	I	I	I	I	I	I	I	I
H	146	**Keep an eye on EU funding calls / opportunities**	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
H	147	**Engage new Member States** to increase annual national contributions and DiSSCo fixed funding	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
M	148	**Provide** training **regarding submitting funding applications and consortium building**	P	P	P	P	I	I	I	I	I	I	I	I				

Discussion and Conclusions

T4.3 has strong dependencies with DPP WP7 (Governance policy and legal framework) and WP8 (Stakeholder engagement & communication strategy):

- · WP7: work on DiSSCo governance, strategy and statutes. National funders commitment is defined and written in the statutes and the way DiSSCo is governed impacts the funders commitment.

- · WP8 and especially T8.1 (National Nodes Engagement): funding could circulate in another way in case there is a specialisation plan decided.

Method

In order to consolidate public revenues (EU and National funders), T4.3 has the following methodology:

1. Benchmark of other ERICs to understand how funding works

2. Modelisation of funding circulation

3. Drafting a calculation method to define Member States' annual contributions to the ERIC and modelisation of its impact

4. Estimation of in-kind requested from Members

5. Synthetic report on EU funding opportunities for an ERIC like DiSSCo

6.3.4 Pre-Commercial Procurement (PCP)

Pre-commercial procurement financial structure

Description   *DPP D4.4*

This task focuses primarily on the procurement framework options needed to support Research & Development. DiSSCo works with both industrial and public partners to co-create services and co-develop products such as, but not limited to, software and digitisation technology and contracts with third parties for goods and services.

The DiSSCo RI requires strategic partnerships with all these actors. A clear strategy allows an organisation to align its procurement process with its long term priorities and objectives, supporting partnership development, scaling up, risk mitigation, efficiency gains and building roadmaps.

By working with an external expert (AcrossLimits), considering how procurement frameworks have been implemented in other relevant European Research Infrastructures and through the internal collaboration with other work packages, we gained insight into the opportunities and challenges that PCP can represent with perspective on other procurement types.

The outcome of this task provides a roadmap for DiSSCo within the PCP framework, leveraging the DiSSCo Strategy developed in DiSSCo Prepare. We do this through offering perspective on funding approaches with their opportunities and challenges for both the DiSSCo Hub as well as its members and funders. We provide an indicative view on the preparation phase to PCP, as this is where the main difference is found. The execution phases follow a defined approach of execution by the external to DiSSCo innovators, and an evaluation by the DiSSCo members at each round.

Timeline and Summary Recommendations

Table 41. Pre-commercial Procurement Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing

Pri Typ	Rec No	Recommendation	2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	149	**Update the DiSSCo Construct development needs (prep needed)**: By listing the development needs according to their industry TRL's and their fit with the 4 strategic pillars defined during DiSSCo Prepare, we can create a list of eligible activities and their priorities, that can then serve as pipeline for PCP and PPI opportunities. a first assesment and selection has been completed, this should be repeated at set intervals (e.g. quarterly), when intermediate perspective is compiled centrally.								P	P	I	I	I	F	F	F	F
H	150	**Establish DiSSCo as legal entity (prep needed)**: In order to leverage the benefits of PCP and PPI during the Construct Phase of DiSSCo, we need a legal entity or one of the consortium partners to assume the procurement role during the construct phase. At which time, also the procurement structure in DiSSCo Hub should be locked.							P	I	F	F						
H	151	**Lock on DiSSCo procurement structure (prep needed)**: the procurement function can be centralised, decentralised, hybrid,.. Each setup has its own merits and drawbacks, nevertheless, a timely choice would benefit the future efforts in DiSSCo.							P	I	F	F						
M	152	**Issue a Prior Information Notice (PIN) (provided the pre-commercial procurement high priority recommendations have been completed)**: with a refined scope of the proposed development, performing a gap analysis to track existing solutions and their TRL levels to define the preferred procurement strategy (PCP, PPI, standard procurement),								P	P	I	I	F	F			

| M | 153 | Work out the PCP and PPI contract notices (provided the pre-commercial procurement high priority recommendations have been completed): Engage in the preparatory work for organising PCP and PPI calls, which includes a clear scope definition, preliminary market consultation leading up to a publication of the contract notice, which details the problem to be solved, its functional requirements and use cases. | | | | | | | | | P | P | I | I | F | F | | |

Discussion and Conclusions

To provide a roadmap for DiSSCo within the EU PCP framework we need a strong connection to the DiSSCo Strategy. We believe having a clear procurement strategy for development is essential to create and strengthen both internal and external DiSSCo partnerships. A clear overall strategy enables us to apply a consistent filter to opportunities for development and procurement, in line with the funders vision.

As an underlying principle, one should consider the benefits of a prior agreed commitment strategy vs a costly punishment strategy, to correspond to having a clear strategy in place proactively supporting/guiding all activities vs a reactive approach.

We recognise that public procurement is a specific undertaking that is governed by legal and regulatory restrictions and obligations. It is important that any procurement team contains or is able to grow expertise and stay updated.

To deliver the expected level of quality (that is to be defined and warranted by the DiSSCo RI), a centralised model is more transparent and straightforward than a distributed model.

While a distributed model would provide all members with a skills development opportunity, the diversity in institutional capacity and staff churn is likely to result in significant fluctuation in the DiSSCo functionality, efficiency and processing time, reflecting on its reputation.

To maximise the impact and output of DiSSCo, dedicated centralised resources appear the straightforward route to achieve the highest quality and build up expertise in favour of the consortium. Nevertheless, dedicating resources is a consideration to be made case by case, especially in the start-up phase of any organisation where it is key to maximise the start-up resources.

A start-up recommendation is to design the function centralised but to delegate the work during the start-up phase to a member organisation that has the specific capability and expertise or to outsource it to a specialist contractor to cover the central dedicated role ad interim.

This assumes one of the members has the expertise available in house and would be temporarily able to provide the service. Temporarily because we assume a rotation of the delegate function between the members, and before dedicating a resource in the hub a few criteria should be met, such as a minimum workload (to justify dedicated resources), and an increased need for internal perspective and expertise ("institutional memory"). The more the DiSSCo goods and services catch on in the market, and a steady stream of revenue and volume is reached, the more a dedicated resource becomes justifiable.

If however, there is no relevant expertise found amongst the members, the investment in central learning and the expertise built up at the Hub could be justified from the start. A lot depends on the terms of the delegated centralised function, how long the appointment lasts, can the level of quality in training be warranted across all members, etc.

Once a legal entity, DiSSCo can start to engage in development that leverages the PCP framework. Therefore, our recommendations to the Construction Master Plan are the following:

- First, to identify PCP and Public Procurement of Innovative Solutions (PPI) expertise within the consortium members' procurement teams.

- Second, to clarify if the funded fixed cost can include a central procurement resource.

- Third, to define concrete executional steps of the chosen approach for readiness by the launch of DiSSCo as a legal entity.

In a parallel survey of business development professionals in the member organisations, dedicated marketing and procurement functions were seen as potential opportunities for efficiency (finance, effort and resource) gains to be made for DiSSCo and within the DiSSCo member community.

Procurement is a key domain for any legal entity. In this report we explore and make recommendations on how buying research and development and innovative products and services could play a key role in improving the efficiency and quality of the future DiSSCo services. The Roadmap for the partnerships project within the EU PCP framework starts at identifying development needs within DiSSCo that exceed the innovative capacity of the consortium. It leads us along the strategic check over the gap analysis via the support requirements at the DiSSCo Hub and the DiSSCo member institutions to the following conclusions.

We have established that among the DiSSCo consortium members there exist unmet innovation needs. A part of those cannot be covered by the innovative capacity present within the consortium, requiring external partnerships. Several of those needs are in line with the overall strategy and would be beneficial for DiSSCo, its members, its funders, its clients and society as a whole, if resolved.

We are convinced that PCP is an opportunity for DiSSCo to advance faster and cheaper on its innovation pipeline. DiSSCo would benefit from designing for a proactive procurement mentality in the organisation during the DiSSCo Construct phase. We believe the ideal upstream procurement person has affinity with business development and a proactive mindset, on top of reactive data processing skills.

From the insights gained during the efforts on task 4.4, we conclude that a centralised procurement approach is more favourable, both from a cost as from a complexity point of view. If the confidence exists that among the members, there is sufficient expertise to manage procurement for DiSSCo, then this should be the approach to start up with. This expertise has to include the processes of pre-commercial procurement and public procurement of innovation. Over time, an evolution towards a dedicated centralised resource can still be made.

However, to meet the quality and reputation expectations of DISSCo, an adequate level of PCP/PPI expertise is a must, which if not found among the members, is best integrated directly at the DiSSCo Hub through a dedicated resource (on the payroll or outsourced). It is with the same reasoning that we advise against distributed models.

Method

We studied the suitability of different procurement strategies (including standard public commercial procurement, PCP and PPI) for DiSSCo, and evaluated how DiSSCo can leverage them.

As a first step, we conducted surveys for procurement awareness within the DiSSCo consortium and to get a view on currently already considered development needs. Details on these surveys are provided in the milestone document MS4.4 (Pijls et al. 2022).

We then upheld the development needs against the following criteria:

- What is the relevance for DiSSCo and does it fit within the DiSSCo strategic programme and construction programme (Hardisty et al. 2020, Alonso & Koureas 2021, Petersen et al. 2022)?

- What is the novelty or TRL of the proposed development need?

- Can the development proposal be executed within DiSSCo by its members, or is there a need for an external partnership?

- Which procurement framework fits best for the given development proposal (PCP, PPI, CP)?

To further detail the strategic relevance per development proposal, we presented the topics for PCP to the CSO. Together with the technical audience in the second all-hands meeting (AHM2 2022), we considered the TRL's and needs for external partnerships. We illustrated which procurement structures are relevant to DiSSCo developments to the fourth Funders Forum (FF4 2022) .

From the CSO, we gathered insight on the 4 DiSSCo strategic pillars, being:

- Digitisation, to accelerate digitisation across European Natural History Collections.

- Access, to create, measure and support multimodal standardised routes to access collections

- Capacity Building, to ensure users and partners are confidently supplying/using data through DiSSCo and are making use of DiSSCo's services

- e-Services development based on user needs defined through engaging with our stakeholders

This, combined with the result of the All Hands Meeting 2 (AHM2) workshop on PCP, helped us to establish that even at the Prepare stage, DiSSCo shows potential for pre-commercial procurement.

We created awareness of these findings at the 4th Funders Forum, after which we compiled our findings and published them in an internal Milestone document (MS4.4) (Pijls et al. 2022).

In parallel, we had access to the expertise of AcrossLimits and procurement examples and analysis in agreement with the European Open Science Cloud (EOSC), which were valuable resources to progress. In this regard DiSSCo and the domain of Biodiversity information is one of the case studies towards the implementation and sustainability of the EOSC in the future (Mergen et al. 2022, Roi et al. 2022). This resulted in a mutual benefit for both sides.

We compiled our learning in an indicative model, based on principle thinking as no reference data was found on the cost and complexity within a consortium during the preparation and execution of a PCP. This model is addressed in the perspective section (Section 5), in which both cost and complexity are projected using assumptions and business logic to see how they vary between the different procurement options for the Hub as well as the members of DiSSCo.

6.3.5 Impact Indicators

Develop indicators of socioeconomic impact

Description   *DPP T1.4*

A set of socioeconomic impact (SEI) indicators was proposed to measure how DiSSCo activities impact: 1) future digitisation of collections, 2) impact research performance and 3) economy at local, national and European levels. The socioeconomic impacts of a RI like DiSSCo, as well as the research and research application activities based on it, are not easily or directly valorised in monetary terms; indicators in use for the cultural sector are comparatively explored. In the context of environmental grand challenges, this WP implements indicators of direct socio-economic impacts related to users, services and areas of impact, such as job creation, technology innovation and more efficient use of existing resources, as well as the impact that DiSSCo based research can have on areas like agriculture, forestry, conservation, bioeconomy, risk management, education and citizen science and broad public engagement.

Timeline and Summary Recommendations

Table 42. Impact Indicators Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
H	154	**Identify detailed data requirements -** Liaise with technical development and partners to provide detailed data requirements on monitoring services, tools, activities to ensure implementation of data gathering and data flow.			I	I	I	I	I	I								
H	155	**Establish a pilot study of indicator quantification** to establish background status of SEI indicators before operation,					I	I	I	I								
M	156	**Define roles within the DISSCo organisation about data gathering, analysis and reporting of SEI indicators.**																

M	157	**Start the design and implementation of the information management system for data gathering, analysis, reporting and archiving**. It might require new developments to implement automatisms.																

Discussion and Conclusions

The definition of a SEI analysis of DiSSCo must have in consideration the type of infrastructure, its phase in the life cycle and, with special focus, the scientific domain of the community and its strategic objectives. The identification of relevant SEI indicators and its implementation is related to:

- **services**: the SEI of the several services to be implemented by DiSSCo should be monitored by indicators that enable it to capture changes and its impact. The planned services are being implemented, therefore is important to interact at an early stage to ensure that data logging of service provides the required details for indicator calculation;

- **users**: the user groups that are served and supported by DiSSCo. These were identified from user cases and user stories compiled by tasks 1.1 and 1.2. A follow up of the impact of these user communities might require surveys to assess the impact that DiSSCo services caused in their activities;

- **areas of impact**: identification of areas related to scientific, industry and innovation, direct and long term socioeconomic impacts, social innovation and other areas where it is expected the impact of DiSSCo activities.

As DiSSCo evolves in its implementation phases, the type and applicability of the indicators are naturally different for each phase, although some are applicable at all phases. This factor needs to be considered also when selecting and implementing indicators for the SEI of DiSSCO.

Method

DiSSCo Prepare WP1, Task 1.4 proposed a set of socio-economic impact (SEI) indicators, based on a review of recent frameworks and studies of socio-economic impact of research infrastructures, the compilation of SEI indicators recommended or used for the assessment of research infrastructures, an assessment of relevance and applicability of socio-economic impact indicators to DiSSCo through a survey to DiSSCo-Prepare partners. It delivered a report on the suggested list of SEI indicators to be used by DiSSCo and guidelines for its implementation. The final list contains approximately 30 indicators[164].

The guidelines document provides orientation on how to implement the quantification of the indicators through a sheet for each indicator, containing detailed information about the indicator name and definition, data requirements, sources and methods for calculation, possible breakdowns, frequency of report and an estimate of the data collection cost. The steps for Its implementation can be:

1. review and identify the specific actors and sources (including tools and interfaces) for gathering data for each indicator;

2. interact with managers of tools/services identified as data sources to ensure mechanisms for data logging and quantification (including breakdown) that support indicators are implemented;

3. design and implement data flows and processing for each indicator, including interfaces with data source systems;

4. identify or design and implement the information management system for the gathering, analysis, reporting and archiving of aggregated information.

Data sources can be single-source (if based on a specific tool or service) or distributed (if dependent on a network of partners), or even dependent surveys to user communities. The defined data flows are, therefore, specific for each indicator.

07 Human Resources

7.1 Overview

Figure 9. Synoptic graph showing the dependencies (arrows) between tasks in this chapter with other tasks or chapters (systems) in DiSSCo Construct.

DiSSCo is anchored in a community already established around the Consortium of European Taxonomic Facilities (CETAF), that gathers major natural sciences collections holders in Europe. This community has grown up over the last few years and DiSSCo now comprises over 170 institutions across 23 countries that host, curate, preserve and give access to over 1.5 Billion physical specimens that jointly represent 80% of the biodiversity described worldwide. The DiSSCo Research Infrastructure aims to serve scientific researchers fundamentally linked to bio- and geodiversity data, but also beyond these specific disciplines to cover other directly linked fields such as health, agriculture, forest, ecology and many others. The multidisciplinary and integrative approach that DiSSCo offers is necessary to address the need for cross-linked information that could effectively underpin the entire research life cycle and would provide open access to mass and precise data.

Scientific institutions and individual researchers are thus at the core of DiSSCo development acting as a two-fold agent: on one hand as providers of data derived from the physical objects held in the facilities and, on the other, as users of the services built on top of that data. However, disparities remain as one of the critical aspects of this community with regard to various aspects, from the size of collections to the level of digitisation of those collections, through the level of skills and competencies of the involved staff, the

harmonisation of procedures and the use of standards and interoperable mechanisms, the integration in national research roadmaps, and in general, the maturity of each institution and their human resources to address the digital transformation that DiSSCo brings to their strategic positioning, not only at the institutional, national and European levels but also internationally.

DiSSCo efforts in terms of Human Resources have been focused in addressing the different aspects of the diverse community integrated in the Research Infrastructure at two levels:

1. Within the community, to identify commonalities and specificities internally, and develop the tools and mechanisms necessary to

   - facilitate smooth and effective operation of the RI by all partners, and

   - provide access to all to exploit DiSSCo services to its full potential.

2. At the DiSSCo hub level, as a central point to operate the DiSSCo Research Infrastructure in a distributed, harmonised and effective manner.

DiSSCo Prepare has included instrumental developments to ensure those two levels are well tackled, from the perspective of the community providing data and also as users of its outcomes.

a) A comprehensive training strategy, to identify gaps of knowledge that will allow the provision of courses and capacity building initiatives to tool up the community to address technological advances, organisational challenges and competency requirements.

b) A Helpdesk that will initiate providers and users in the concept of digital objects and the DiSSCo architecture and will help the community navigate through the complex landscape that surrounds digital access to FAIR data.

c) A policy to support Human Resources management, including actions such as secondments, that will activate distributed operation of the RI and will facilitate the composition of a team that will lead the successful operation of DiSSCo in an integrated system.

d) A Self-Assessment tool that will allow individuals to evaluate their level of maturity in the digital realm towards the implementation of the Infrastructure within their working procedures and objectives.

e) A work-flow to set-up linkages to the different tools created under DiSSCo Prepare and/or to be developed during the transition/construction phases of the project, among others,

   i)   Knowledgebase

   ii) Specialisation Plan

iii)   <u>Dashboards</u> for experts distribution.

7.2 Timeline and Summary Recommendations

Table 43. Human Resources Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

				"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing																
	Pri	Rec		2023				2024				2025				2026			
Task Name	Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
7.3.2 Training Strategy	Tr	165	The Training Strategy for DiSSCo constitutes the starting point to first develop and then implement the DiSSCo Training Programme. It provides substantiated information about the existing capacities and the detected demand across the community that needs to be further elaborated. During the Transition phase, specific actions might be required: **Embed the DiSSCo Training into the CETAF-DEST and articulate how to implement it, following the recommendations from DPP.**		P	P	I	F											
7.3.2 Training Strategy	Tr	166	During the Transition phase, specific actions might be required: **Deploy the DiSSCo training service into annual training programmes in accordance to the priorities**				P	I	F										

			detected to ensure DiSSCo operation across institutions as providers of data.																
7.3.2 Training Strategy	Tr	167	During the Transition phase, specific actions might be required: **Align the DiSSCo programmes with other complementary resources to avoid duplication of efforts (such as GBIF) by subscribing specific agreements for collaboration.**		P	P	P	I	F										
7.3.1 Human Resources Policy	H	158	**Hire a HR manager to implement human resources management actions based on the policy defined.**	P	P	P	I	I	I	I	F								
7.3.1 Human Resources Policy	H	159	**Identify the number and the profiles of professionals needed for DiSSCo implementation.**	P	P	P	I	I	I	I	F								
7.3.1 Human Resources Policy	M	160	**Define a compensation plan that can be attractive to the selected professional.** The figures will be necessary for DiSSCo implementation.	P	P	P	P	I	I	I	I	I	I	I	F				
7.3.1 Human Resources Policy	M	161	**Define a recruitment strategy and design a professional training program to acquire the skills required by DiSSCo implementation.** Training should go beyond job-related skills development like project management, IT development, open science, law, EU projects or service delivery, and include skills focusing on long-term personal development such as leadership, communication, networking skills, stakeholder engagement, project	P	P	P	P	I	I	I	I	I	I	I	F				

			management, presentation skills, facilitation skills, IT and informatics skills and so on.																
7.3.1 Human Resources Policy	M	162	**Write specific guidelines for mobility, work conditions and equality of treatment.**	P	P	P	P	I	I	I	I	I	I	I	F					
7.3.3 Capacity Self-Assessment (Digital Skills including Dashboard)	M	171	**Ensure software development for digital maturity self-assessment tool:** The digital maturity self-assessment tool is being developed under T7.3 but will require ongoing support and refinement including digital maturity content. Please refer to T7.3 for recommendations on future development needs for this tool.																	
7.3.3 Capacity Self-Assessment (Digital Skills including Dashboard)	M	172	**Define questions for sub-categories of digital maturity and self-assessment tools:** MS3.4 details the high level content and subcategories for the digital maturity tool, and examples of content for the self-assessment, but further work will be required to define questions in each sub-category. This should link closely with outputs from WP2, to ensure the self-assessment tool can support institutions and individuals by linking to available training and resources.																	
7.3.3 Capacity Self-Assessment (Digital	M	173	**Ensure role/job profiles are setup in the Knowledgebase (KB):** This task recommends that relevant role/job profiles are set up in the DiSSCo KB. This work is being progressed in WP2.																	

Section	L	#	Description												
Skills including Dashboard)															
7.3.1 Human Resources Policy	L	163	**Design a <u>training</u> program on multidisciplinary teamwork to enhance the effectiveness of teams.**	P	P	P	P	I	I	I	I	I	I	I	F
7.3.1 Human Resources Policy	L	164	With the purpose of ensuring compliance with high quality standards in the exercise of DiSSCo's functions **it is relevant to carry out systematic assessments/audits** in order to identify and correct non-conformities, as well as the creation of dedicated institutional communication channels in order to allow the dissemination of clear, concise and useful information about DiSSCo.	P	P	P	P	I	I	I	I	I	I	I	F
7.3.2 Training Strategy		168	During the construction phase and towards practical implementation of training, other activities shall be supported: **Integrate DiSSCo online training into the e-learning platform provided by TETTRIs project.**		P	P	I	I	I	I	I	I	I	I	F
7.3.2 Training Strategy		169	During the construction phase and towards practical implementation of training, other activities shall be supported: **Link training provision with expertise availability by creating a comprehensive lively web-based database of experts in different disciplines and fields necessary for digitisation work.**				P	P	P	I	I	I	I	I	F

			During the construction phase...															

7.3.2 Training Strategy	170	During the construction phase and towards practical implementation of training, other activities shall be supported: **Investigate recognition support with academia.**				P	P	P	P	I	I	I	F				
7.3.4 Distributed Teams and Secondment	174	**Continue to evaluate distributed ways of working throughout construction.** The distributed working recommendations will be relevant throughout the construction phase of DiSSCo.															
7.3.4 Distributed Teams and Secondment	175	**Continue to evaluate secondment (tech infrastructure developers) to CSO during construction.** The secondment recommendations contained within MS3.9 may be relevant during the construct phase of DiSSCo, depending on whether secondment to the DiSSCo Coordination and Support Office (CSO) is seen as beneficial during this period (e.g. for development of the technical infrastructure). However, distributed team-working alongside existing roles may be a more appropriate model. There are no current construction tasks as no decision has been taken to implement a secondment model.															

7.3 Task Summaries

7.3.1 Human Resources Policy

Description   *DPP D2.3*

This task establishes a Human Resources Policy (HRP) that allows the implementation of DiSSCo strategy and operational objectives while ensuring a scientific, technologically and socially stimulating working environment. It builds on the strategic map that establishes the alignment between the HRP and the strategic objectives of DiSSCo. This HRP requires the identification of factors that induce strategic performance in Human Resources Management (HRM), which contributes in turn to the successful implementation of the DiSSCo mission. Work includes a landscape analysis of HRP at European Strategy Forum on Research Infrastructures (ESFRI), European Research Infrastructure Consortium (ERIC) and national level and analysis of their suitability to the entity model proposed by Task 7.2.

This requires interviews with partner's human resources managers to identify cross-sectional human resources practices that can be generalised. A portfolio of roles and competencies necessary to support DiSSCo operation is defined in close collaboration with Task 3.1. These include digital roles such as research data curator, digitiser, data architect, business analyst, and programme/project managers, and management roles. These need to be developed to recognise the necessity of hybrid roles in smaller organisations that do not have the resources to dedicate full positions to such specialised tasks.

The HRP covers both current roles in DiSSCo facilities and standardised role descriptions from external organisations where relevant. Also, it defines human resources practices such as training models, performance management and team-based organisation. Benchmarking, mentoring, training and mobility are pivotal aspects for permanent staff capacity enhancement and personal development.

Timeline and Summary Recommendations

Table 44. Human Resources Policy Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
H	158	**Hire a HR manager to implement human resources management actions based on the policy defined.**	P	P	P	I	I	I	I	F									
H	159	**Identify the number and the profiles of professionals needed for DiSSCo implementation.**	P	P	P	I	I	I	I	F									
M	160	**Define a compensation plan that can be attractive to the selected professional.** The figures will be necessary for DiSSCo implementation.	P	P	P	P	I	I	I	I	I	I	I	F					
M	161	**Define a recruitment strategy and design a professional <u>training</u> program to acquire the skills required by DiSSCo implementation.** Training should go beyond job-related skills development like project management, IT development, open science, law, EU projects or service delivery, and include skills focusing on long-term personal development such as leadership, communication, networking skills, stakeholder engagement, project management, presentation skills, facilitation skills, IT and informatics skills and so on.	P	P	P	P	I	I	I	I	I	I	I	F					
M	162	**Write specific guidelines for mobility, work conditions and equality of treatment.**	P	P	P	P	I	I	I	I	I	I	I	F					
L	163	**Design a <u>training</u> program on multidisciplinary teamwork to enhance the effectiveness of teams.**	P	P	P	P	I	I	I	I	I	I	I	F					
L	164	With the purpose of ensuring compliance with high quality standards in the exercise of DiSSCo's functions **it is relevant to carry out systematic assessments/audits** in order to identify and correct non-conformities, as well as the creation of dedicated institutional communication channels in order to allow the dissemination of clear, concise and useful information about DiSSCo.	P	P	P	P	I	I	I	I	I	I	I	F					

Discussion and Conclusions

Four critical success factors (CSF) considered essential for HRP were identified: (1) engaging people to achieve organisational objective, (2) establish the right human resources capabilities and competencies; (3) flat management structure, flexibility, autonomy and dialogue; (4) improve human resources operational excellence; (5) measure impact of human resources products and services.

Each CSF has an associated set of HRM practices that promote the corresponding factor:

- Engaging people to achieve the goals requires jobs designed to promote skill variety and opportunities to learn through complex and autonomous tasks and diverse working experiences. Also, it requires conditions to develop a working climate of social support and supportive supervision.

- Establishing right human resources capabilities and competencies requires recruitment and selection procedures which are open, efficient, transparent, supportive and internationally comparable, as well as tailored to the type of positions advertised. Also, it requires training practices developed around the key competency cluster recommended by DPP WP3 (Capacity enhancement) on DiSSCo Competencies and Digital Maturity of which we would emphasise leadership and management of people, resources and change; communication, collaboration, and training others; digital and technological skills; data and information skills; collections management and curation. Finally, the creation of opportunities for career development is also a desirable HRM practice.

- Improving human resources operational excellence is achieved through flexibility of working conditions from part-time work to remote working, policies related to ethics, equal opportunity and equity issues, and mobility between roles, internal and external.

- Measuring performance and impact requires the implementation of a performance management system that makes possible gathering the information needed to build the various programmes (promotion, remuneration, training, etc.) and supporting decisions on the career advancements of staff members. In addition, the feedback of the evaluation should allow the employee to improve their work, autonomy, and competence, as well as being aware of the opportunities offered by the organisation and to express their achievements and their wishes for development, training and career, as well.

Finally, the HRP proposed here is, in the first instance, designed to the DiSSCO hub structure because of its central coordinating role among the various stakeholders, but also because the various national nodes are subject to the personnel policy of their domestic organisation.

The implementation of the HRP is dependent upon the definition of DiSSCo's organisational structure as well as the number of effective members. The implementation of a HRM system is a longitudinal process where practices such as recruitment and selection come at an early stage and performance management, for instance, makes more sense in a later stage of DiSSCo Hub life.

Method

One sub-task was a landscape analysis of HRP across European Research Infrastructures, which also operate as distributed RIs like DiSSCo. Six of these RIs accepted our invitation to participate in our study (CLARIN, DARIAH, EMBL, EPOS, LifeWatch and GBIF).

This analysis permitted us to identify general HR practices that can be used as benchmarking to develop DiSSCo people management policy. In order to identify the strategies and practices considered relevant in RIs HRP a survey and analysis of the RIs available HRM documents was carried out (DiSSCo Prepare WP2 – Ms2.5[165]).

A second sub-task was to build a portfolio of roles and competencies necessary to support DiSSCo operation. That was done by consulting DiSSCo partners and the six RIs that participated in the survey.

In order to select the best HR practices and guidelines we used the landscape analysis of HRP across European RIs produced earlier by WP2 along with relevant scientific literature on HRM.

7.3.2 Training Strategy

Description   *DPP D2.1*

This task is built upon information already gathered in previously and currently running projects, in addition to information collected by Task 3.1-Capacity enhancement to identify critical competencies for digitisation of collections. The ultimate aim of the task is to develop a training strategy with distinct channels and modes of accessing training to address the identified needs of the DiSSCo users and providers of data. Therefore, in such a framework, personnel capacity building is considered both on the side of the data suppliers (collection managers, data digitisers, museum technicians, etc.) and also from the perspective of the users of DiSSCo (researchers, policy makers, students, environmental Agencies, NGOs, etc.).

The training scheme refines and complements existing training programmes organised within the DiSSCo community (see Table 4 - objs. d, i and r; CETAF DEST (Distributed

European School of Taxonomy), Mobilising Data, Policies and Experts in Scientific Collections (MOBILISE, COST Action) and BIOTALENT (Erasmus+ funded project), and considers both academic and continuous professional training. This task also develops the recommendations produced by the DiSSCo ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage) Design Project and elaborates, among others, paths to integrate bioinformatics-related topics into formal education, while formalising collaborative actions with e.g. academics, professional associations and providers of vocational training, etc. The definitions of DiSSCo training programmes are based on the 'train the trainer' principle, aiming at creating a strong network of trainers able to provide support to their local communities, both institutionally and nationally. Additionally, diversity based on aspects such as educational resources, cultural differences and policies and regulations applicable, have also been considered.

Timeline and Summary Recommendations

Table 45. Training Strategy Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

Pri Typ	Rec No	Recommendation	"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Tr	165	The Training Strategy for DiSSCo constitutes the starting point to first develop and then implement the DiSSCo Training Programme. It provides substantiated information about the existing capacities and the detected demand across the community that needs to be further elaborated. During the Transition phase, specific actions might be required: **Embed the DiSSCo Training into the CETAF-DEST and articulate how to implement it, following the recommendations from DPP.**		P	P	I	F											
Tr	166	During the Transition phase, specific actions might be required: **Deploy the DiSSCo training service into annual training programmes in accordance to the priorities detected to ensure DiSSCo operation across institutions as providers of data.**				P	I	F										

	#	Action																		
Tr	167	During the Transition phase, specific actions might be required: **Align the DiSSCo programmes with other complementary resources to avoid duplication of efforts (such as GBIF) by subscribing specific agreements for collaboration.**	P	P	P	I	F													
	168	During the construction phase and towards practical implementation of training, other activities shall be supported: **Integrate DiSSCo online training into the e-learning platform provided by TETTRIs project.**	P	P	I	I	I	I	I	I	I	I	F							
	169	During the construction phase and towards practical implementation of training, other activities shall be supported: **Link training provision with expertise availability by creating a comprehensive lively web-based database of experts in different disciplines and fields necessary for digitisation work.**				P	P	P	I	I	I	I	I	F						
	170	During the construction phase and towards practical implementation of training, other activities shall be supported: **Investigate recognition support with academia.**				P	P	P	P	I	I	I	F							

Discussion and Conclusions

From the combination of the Training Catalogue (from SYNTHESYS+) and the Training Needs (from DPP), it has been possible to detect which is the level of training capacities/requirements in each of the five dimensions of DiSSCo RI. While the scientific aspects are well covered, the digital transformation that DiSSCo encompasses is still far from being achieved and training is strongly required to tool-up the institutions' staff to be able first to provide meaningful data in the adequate format and secondly, make use of the RI for supporting their research.

Data and technological aspects require specific training to enhance proficiency and allow digitisation related personnel to curate and manage in a standardised format the data provision. Capacity building is equally needed to support all digitisation phases, from imaging to metadata aggregation. On the other side, financial and governance dimensions, which initially were not considered under the scope of this personnel, become increasingly

part of the multidisciplinary tasks that technicians, curators, collections managers and heads of departments need to perform.

To convey all these needs under a unique training programme has been considered a challenge and a modular scheme has been proposed to tackle training in different axes, horizontally as to cover the different competencies needed within an institution, as well as vertically, to achieve the adequate degree of proficiency. Furthermore, career paths and professional recognition are considered to be at the core of the training programme, to ensure sustainability and expansion of the personnel base needed to operate DiSSCo and provide efficient and meaningful DiSSCo services. Additionally, diversity of the DiSSCo community is at the basis of specific needs such as multilingual facilities that the DiSSCo training programme needs to cover, as well as the provision of an efficient, user-friendly, sustainable and comprehensive platform where to allocate and display the resources produced for training. Further information, concluding remarks and recommendations are in the final product D2.1 Training strategy report[166].

Method

The work towards building a Training Strategy for DiSSCo has included reviewing existing best practices for training delivery, analysing reports and recommendations produced under other DiSSCo-linked initiatives (as mentioned above) and creating a baseline of the needs and expectations for training on digitisation work on which to anchor the definition of the strategy, and envisage how to better implement it overtime. An additional search for the most suitable platforms to host such a training programme has been run. Similarly, governance options to operate the training programme have been proposed, since this is considered a pivotal capacity service for the DiSSCo community.

Specific activities have included: a) compilation of needs for skills/competencies; b) identification of training providers/platforms; c) landscape analysis of best practices for training delivery; and d) integration of all training strategy elements, to jointly provide a final report with a recommendation for setting up the DiSSCo Training Strategy. Building on top of the Training Catalogue produced under SYNTHESYS+, which was considered the reference for the "offer" side, a survey was launched to obtain feedback from DiSSCo members on the "demand" side and thus to have an overarching view on how to deploy training for increasing skills and facilitating acquisition of new competencies across DiSSCo community, and ultimately to enable exploitation of the RI to its full potential. Further analysis, with reports on opportunities and obstacles was run and the final report was produced with concluding remarks and recommendations.

7.3.3 Capacity Self-Assessment (Digital Skills including Dashboard)

Improve digital skills and competencies across DiSSCo facilities

Task 3.1 addressed one of the key aims of this WP - to describe the mechanisms and tools to improve digital skills and competencies across facilities, in a way that can be scaled to different sizes and types of organisation within the DiSSCo consortium, recognising that in many cases, there is a need for hybrid roles encompassing a mixture of digital skills and competencies by a single individual, especially in smaller organisations.

While this Task was originally expected to develop a competency framework for DiSSCo, analysis of existing frameworks suggested both that this would represent a duplication of existing efforts, and that there are substantial barriers to use (e.g. institutional practices or national employment law) which would prevent such a framework being effective. Similarly, analysis of sources of data about individual competency (such as LinkedIn) identified that there are no existing sources that can usefully be incorporated into a dedicated DiSSCo dashboard at this point.

Instead, therefore, the Task focused on providing the insights from these analyses, including identifying key competency 'clusters' for DiSSCo that have been used to inform other Tasks in WP2 (Human Resources, Training & Users Support) and WP4 (Business Framework). In addition, the Task developed functional and non-functional requirements and first content for a DiSSCo Digital Maturity Self-Assessment tool. This tool is intended to help institutions and teams self-assess key areas of their capability and capacity, to prioritise areas for enhancement; and to help the DiSSCo central organisation target training, support and policies/processes such as secondment, to maximise capacity for DiSSCo delivery.

There were no development resources allocated to Task 3.1, however working with Task 7.3 Develop and establish DiSSCo policies, we identified scope for the initial build of this tool to be undertaken alongside the development of the DiSSCo Policy tool, ensuring a consistent platform and approach.

Timeline and Summary Recommendations

Table 46. Capacity Self-Assessment (Digital Skills including Dashboard) Recommendations

NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No

			"P" for "Planning", "I" for "Implementation, "F" for "Finalisation", "O" for Ongoing														
Pri	**Rec**		2023				2024				2025				2026			
Typ	**No**	**Recommendation**	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

M	171	**Ensure software development for digital maturity self-assessment tool:** The digital maturity self-assessment tool is being developed under T7.3 but will require ongoing support and refinement including digital maturity content. Please refer to T7.3 for recommendations on future development needs for this tool.																		
M	172	**Define questions for sub-categories of digital maturity and self-assessment tools:** MS3.4 details the high level content and subcategories for the digital maturity tool, and examples of content for the self-assessment, but further work will be required to define questions in each sub-category. This should link closely with outputs from WP2, to ensure the self-assessment tool can support institutions and individuals by linking to available training and resources.																		
M	173	**Ensure role/job profiles are setup in the Knowledgebase (KB):** This task recommends that relevant role/job profiles are set up in the DiSSCo KB. This work is being progressed in WP2.																		

Discussion and Conclusions

The analyses undertaken for MS3.1[167] and MS3.3[168], as well as discussions at the DiSSCo All Hands Meeting in January 2021, indicated that a new DiSSCo competency framework was unlikely to be effective, owing to the variety of existing resources available and the low take up of these across different institutional contexts. Instead, the task focussed on the development of a DiSSCo Digital Maturity Self-Assessment tool, which would allow teams and organisations to self-assess and identify areas for improvement. The task created a high level design blueprint for the tool, which was further developed in T7.3, and also provided a detailed description of the content for this self-assessment.

The analysis of roles and competencies from this task has been used in T4.1 Cost Book to define standard staff seniority levels for the DiSSCo Cost Book, and has been used as input into the training strategy in T2.1.

A prototype of the digital maturity tool was created during DiSSCo Prepare phase as part of T7.3. The further use of this platform and further development and deployment of the tool are dependent on the wider planning and progress of DiSSCo e.g. the training strategy.

Method

MS3.1 and MS3.3 examined insights from previous DiSSCo-linked projects; evaluated competency and skills frameworks; analysed sources of competency and capability data; and assessed several tools and approaches to assess organisational digital capabilities and maturity. A set of functional and non-functional requirements for a digital maturity tool were captured, as well as initial tool content being drafted, and this tool was further developed under T7.3.

7.3.4 Distributed Teams and Secondment

Develop DiSSCo secondment and distributed team working practices

Description   *DPP D3.3*

Implementing the vision for DiSSCo requires improving staff capacity and sharing expertise across partner institutions to develop the infrastructure. This challenge is most acute in technical, engineering and project management areas where specialist staff are fundamental to delivering core parts of the DiSSCo Programme, and where expertise may be scarce and is likely to be distributed across different nations and institutions.

Building capacity and developing successful distributed teams may require a DiSSCo-supported secondment approach for the fixed-term assignment of staff to other teams, and/or tools and approaches to support productive distributed teams working for dedicated cross-institutional distributed teams. This is the key to sharing best practices and developing trust across the consortium.

This task therefore explored secondment models and distributed team working tools and practices.

Timeline and Summary Recommendations

Table 47. Distributed Teams and Secondment Recommendations
NB: Table ordered by Priority Type (Transition is highest, then H, M, L); then by Rec No.

Pri	Rec		"P" for "Planning", "I" for "Implementation, "F" for Finalisation", "O" for Ongoing															
			2023				2024				2025				2026			
Typ	No	Recommendation	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
	174	**Continue to evaluate distributed ways of working throughout construction.** The distributed working recommendations will be relevant throughout the construction phase of DiSSCo.																
	175	**Continue to evaluate secondment (tech infrastructure developers) to CSO during construction.** The secondment recommendations contained within MS3.9 may be relevant during the construct phase of DiSSCo, depending on whether secondment to the DiSSCo Coordination and Support Office (CSO) is seen as beneficial during this period (e.g. for development of the technical infrastructure). However, distributed team-working alongside existing roles may be a more appropriate model. There are no current construction tasks as no decision has been taken to implement a secondment model.																

Discussion and Conclusions

Secondment: a DiSSCo secondment policy should benefit not only the individuals and institutions involved, but also the wider needs of the DiSSCo infrastructure. It should deliver skills growth or transfer needed for DiSSCo and/or delivery of DiSSCo components. This means that if such a scheme is felt to be needed / useful to DiSSCo it needs to be at least part-funded and incentivised centrally. In many cases, however, distributed team working models may be more flexible and appropriate. More detail is available in MS3.9.

Distributed Working: Interviews about distributed working tools and practices, as well as a summary of approaches taken in the distributed working pilot (development of DiSSCo policy and digital maturity tools in T7.3), show that key benefits of changes to working practices during and after the Covid 19 pandemic are the widespread adoption of distributed working tools; the greater choice and flexibility available to individuals about where and how they work; and the greater accessibility of certain events such as online conferences. On the other hand, face to face contact is still highly valuable and likely to be

particularly important at the set-up or other key stages of distributed team working. As well as technical skills, project management and business input are important to infrastructure development in distributed teams. Wherever possible, DiSSCo should allow teams flexibility to decide the tools and approaches that work for them, within an overall framework of advice and options to ensure coherent approaches. More detail is available in D3.3 (incorporating MS3.10).

Method

Secondment: MS3.9 examines examples of secondment schemes and suggests an approach to secondment for DiSSCo.

Distributed Working: D3.3 (which incorporates MS3.10) summarises the insights from a series of semi-structured interviews with key roles in DiSSCo including developers, CSO staff and digitisation teams. These interviews reflect the widespread adoption of distributed working tools following the Covid Pandemic, and other changes to working practices. It also summarises processes and learnings from a pilot of distributed working to deliver DiSSCo functionality - development of the policy and digital maturity self-assessment tools under Task 7.3 (Develop and establish DiSSCo policies) was selected as this pilot.

08 Appendix

8.1 DiSSCo Prepare Project Deliverables (links)

Please note:

- · *DOIs will be activated once deliverables have been approved and will redirect to DiSSCo's Knowledgebase.*

- · *Deliverables in boldfaced type below have corresponding Task Summaries in the Construction Master Plan.*

D1.1 Report on life sciences use cases and user stories

https://doi.org/10.34960/xhxw-cb79

D1.2 Report on Earth sciences use cases and user stories

https://doi.org/10.34960/n3dk-ds60

D1.3 Report on relevant criteria for prioritisation of the digitisation.

https://doi.org/10.34960/nmhf-wf81

D1.4 Report on socioeconomic impact indicators of DiSSCo and DiSSCo-enabled research and research applications

https://doi.org/10.34960/zg92-j758

D2.1 Training Strategy Report

https://doi.org/10.34960/zm3z-s230

D2.2 Recommendations on the Helpdesk and user support services

https://doi.org/10.34960/5aqf-d131

D2.3 Human Resources Policy

https://doi.org/10.34960/e89g-2j14

D3.1 Summary Insights and Recommendations on DiSSCo Competencies and Digital Maturity

https://doi.org/10.34960/3pc3-pp32

D3.2 DiSSCo Digitisation Guides Website - Consolidating Knowledge on Collections Mobilisation

https://doi.org/10.34960/rh2m-c502

D3.3 Distributed Teamworking Practices and Pilot

https://doi.org/10.34960/yc9r-s177

D4.1 The Cost Book for DiSSCo

https://doi.org/10.34960/kdkr-sf06

D4.2 Cost Book sheet for all the DiSSCo services

https://doi.org/10.34960/vk0e-6t03

D4.3 Report on recommendations for the most suitable model

https://doi.org/10.34960/55r6-bn19

D4.4 Roadmap for the partnerships project within the EU PCP framework

https://doi.org/10.34960/vf2c-6g52

D4.5 Models for government funding

https://doi.org/10.34960/qy5r-3p89

D5.1 DiSSCo Knowledgebase for technical development

https://doi.org/10.34960/myr2-vm35

D5.2 DiSSCo Modelling Framework

https://doi.org/10.34960/e3nv-zh69

D5.3 DiSSCo Digital Specimen Object Specifications

https://doi.org/10.34960/vn64-ws93

D5.4 A best practice guide for semantic enhancement and improvement of semantic interoperability

https://doi.org/10.34960/ajxs-zr25

D5.5 Construction plans for the improvement of technical infrastructure in the areas of geo-collection data and taxonomic services

https://doi.org/10.34960/dzs0-xa94

D6.1 Harmonization and migration plan for the integration of CMSs into the coherent DiSSCo Research Infrastructure

https://doi.org/10.34960/366d-sf49

D6.2 Implementation and construction plan of the DiSSCo core architecture

https://doi.org/10.34960/50b9-kj05

D6.3 A generalised set of API specifications for interaction with the DiSSCo core architecture

https://doi.org/10.34960/htzv-nw73

D6.4 Implementation of the DiSSCo Data Management Plan (DMP) and ENVRI FAIR compliance of DiSSCo data services

https://doi.org/10.34960/1jqv-1335

D7.1 A detailed plan for governance structure and function, and participation framework

https://doi.org/10.34960/j2zd-sn88

D7.2 Draft statutes and by-laws; implementation plan

https://doi.org/10.34960/ft1c-qz88

D7.3 Assessment tools and direction map to the implementation of common DiSSCo policies

https://doi.org/10.34960/c80z-7152

D8.1 Communication and Dissemination Strategy

https://doi.org/10.34960/qk02-4641

D8.2 Thematic specialisation plan

https://doi.org/10.34960/bp9d-5c57

D8.3 Partnership Best Practices

https://doi.org/10.34960/sh1t-4541

D9.1 Kick-Off Meeting

https://doi.org/10.34960/2egp-j862

D9.2 Data Management Plan

https://doi.org/10.34960/8mzj-g791

D9.3 1st DiSSCo Prepare All Hands Meeting

https://doi.org/10.34960/9n9m-9424

D9.4 2nd DiSSCo Prepare All-Hands meeting

https://doi.org/10.34960/zcq4-sh54

D9.5 Final Project Conference

https://doi.org/10.34960/ar8f-h849

D9.6 Compilation of Construction Masterplan

https://doi.org/10.34960/cy1m-b238

8.2 Acronyms

AAI: Authentication and Authorization Infrastructure
AARC: Authentication and Authorisation for Research and Collaboration
ABCD: Access to Biological Collection Data
AHM: All Hands Meeting
AI: Artificial Intelligence
AISBL: International Association of the European Heritage Network
API: Application Programming Interface
ARISE: Authoritative and Rapid Identification System for Essential biodiversity information
BBMRI: Biobanking and BioMolecular Resources Research Infrastructure
BGBM: Botanischer Garten und Botanisches Museum Berlin
BGE: Biodiversity Genomics Europe
BiCIKL: Biodiversity Community Integrated Knowledge Library
BioCASe: Biological Collection Access Service
Bio DT: Biodiversity Digital Twin
BOLD: Barcode of Life Data System

BPMN: Business process model and notation

CDD: Collection Digitisation Dashboard

CDM: Clean Development Mechanism

CETAF: Consortium of European Taxonomic Facilities

CETAF-DEST: CETAF-Distributed European School of Taxonomy

CETAF ISTC: CETAF - Information Science and Technology Committee

CLARIN: Common Language Resources and Technology Infrastructure

CMP: Construction Master Plan

CMS: Collection Management System

COL: Catalogue of Life

CSF: Critical Success Factors

CSIC: Agencia Estatal Consejo Superior de Investigaciones Científicas

CSO: Coordination and Support Office

CWL: Common Workflow Language

D: Deliverable

DD: DiSSCo Development

DARIAH: Digital Research Infrastructure for the Arts and Humanities

DCOLL: German Natural Sciences Collections as an Integrated Research Infrastructure

DES: Digital Extended Specimen

DestinE: EU's Destination Earth

DevOps: Development and Operations

DiCSN: DiSSCo Collection Management System Support Network

DINA: Digital information system for Natural history data

DiSSCo: Distributed System of Scientific Collections

DMF: DiSSCo Modelling Framework

DMP: Data Management Plan

DOA: Digital Object Architecture

DOI: Digital Object Identifiers

DOIP: Digital Object Interface Protocol

DPP: DiSSCo Prepare Project

DS: Digital Specimen

DSArch: Digital Specimen Architecture

DSOS: Digital Specimen Object Specifications

DSR: Digital Specimen Repository

DwC: Darwin Core

ECSA: European Citizen Science Association

EEA: European Economic Area

EEIG: European Economic Interest Grouping

EGTC: European Groupings of Territorial Cooperation

ELI: Extreme Light Infrastructure

ELIXIR: European life science infrastructure for biological information

eLTER: Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research

ELViS: European Loans and Visit System

EMBL-EBI: European Bioinformatics Institute

ENA: European Nucleotide Archive

ENVRI: Environmental Research Infrastructures

EOL-Trait Bank: Encyclopedia of Life-Trait Bank

EOSC: European Open Science Cloud

EPOS: European Plate Observing System

EPOS DCAT-AP European Plate Observing System Data Catalog Vocabulary-Application

ERA: European Research Area

ERIC: European Research Infrastructure Consortium

E-RIHS: European Research Infrastructure for Heritage Science

ESFRI: European Strategy Forum on Research Infrastructures

ETL: Extract, Transform and Load

EU: European Union

FAIR: Findable, Accessable, Interoperable, and Reusable

FAIRDO: FAIR Digital Object

FAQs: Frequently Asked Questions;

FDO: FAIR Digital Object

FDOF: FAIR Digital Object Forum

FF: Funders Forum

GA: General Assembly

GBIF: Global Biodiversity Information Facility

GDPR: General Data Protection Regulation

GeoCASe: Geoscience Collections Access Service

GEOSS: Global Earth Observation System of Systems

GFBio: German Federation for Biological Data

GGBN: Global Genome Biodiversity Network

GRNet: Greek Research and Technology Network

GRScicoll: The Global Registry of Scientific Collections

HNHM: Hungarian Natural History Museum

HR: Human Resources

HRM: Human Resources Management

HRP: Human Resources Policy

IAM: Identity and Access Management

IBER-BAS: Institute of Biodiversity and Ecosystem Research – Bulgarian Academy of Sciences

IBSAS: Plant Science and Biodiversity Centre, Slovak Academy of Sciences

ICEDIG: Innovation and consolidation for large scale digitisation of natural heritage

ID: Digital Identity

iDigBio: Integrated Digitized Biocollections

IdP: Identity Providers

iGA: interim General Assembly

INSDC: International Nucleotide Sequence Database Collaboration

IO: International/Intergovernmental Organisation

IRLs: Implementation Readiness Levels

IT: information technology

JASC: Joint Academic Coding System

JSON: JavaScript Object Notation

KB: Knowledgebase

KPIs: Key Performance Indicators

LUOMUS: Finnish Museum of Natural History

maDMP: machine-actionable Data Management Plan

MfN: Museum für Naturkunde Berlin – Leibniz Institute for Evolution and Biodiversity Science

MHNHL: National Museum of Natural History Luxembourg

MIDS: Minimum Information about a Digital Specimen

MLaaS: Machine Learning as a Service

MNHN: Muséum National d'Histoire Naturelle

MOBILISE: Mobilising Data, Policies and Experts in Scientific Collections

MoSCoW: Must have, Should have, Could have, Won't have

MS: Milestone

MVP: Minimum Viable Product

NA: Network Activities

Naturalis: Naturalis Biodiversity Center

NH: Natural History

NHM: Natural History Museum, London

NHM-UIO: Natural History Museum, Oslo University

NHMW: Natural History Museum Vienna

NM: Národní muzeum

NNs: National Nodes

NRM: Swedish Museum of Natural History

NSCs: Natural Science Collections

ODRL: Open Digital Rights Language

openDS: open Digital Specimen

ORCID: Open Researcher and Contributor ID

OWASP: Open Web Application Security Project

RBGE: Royal Botanic Garden Edinburgh

PID: Persistent identifier

PCP: Pre-commercial Procurement

PIN: Prior Information Notice

PPI: Public Procurement of Innovation

PROV-O: Provenance Ontology

RA: Registration Agency

RBINS: Royal Belgian Institute of Natural Sciences

RDA: Research Data Alliance

RDM: Research Data Management

RECOLNAT: Réseau national des collections naturalistes

REST: Representational State Transfer

RI: Research Infrastructure

ROR: Research Organization Registry

SAB: Scientific Advisory Body

SciColl: Scientific Collections International

SDR: Specimen Data Refinery

SEI: Socio Economic Impact

SEMAF: Flexible Semantic Mapping Framework

SGN: Senckenberg Gesellschaft für Naturforschung

SLA: Service Level Agreement

SOP: Standard Operating Procedures

SRL: Service Readiness level

SWG: Statutes Writing Group

SWOT: Strengths, Weaknesses, Opportunities and Threats

SYNTHESYS: Synthesis of systematic resources

TA: Transnational Access

TAB: Technical Advisory Board

Tartu: University of Tartu

TDWG: Biodiversity Information Standards, formerly

Taxonomic Databases Working Group

TETTRIs: Transforming European Taxonomy through Training,

Research, and Innovations

TRL: Technical Readiness Level

TS: Terminology Services

UCAS: Unified Curation and Annotation System

UCPH: University of Copenhagen

UI/UX: User Interface/User Experience

UniFi: University of Florence - Natural History Museum

UoC-NHMC: Natural History Museum of Crete, University of

Crete

UPORTO: Natural History and Science Museum of the University of Porto

UTARTU: University of Tartu

UW: University of Warsaw

VA: Virtual Access

VAT: Visualization, Analysis and Transformation System

WAI: Web Accessibility Initiative

WP: Work Package

W3C-DCAT: World Wide Web Consortium-Data Catalog Vocabulary

8.3 Endnotes

In the following endnotes, lower case letters indicate that the citation is repeated in another location in the text.

1. ICEDIG a, b, c, d: https://www.dissco.eu/icedig/ , https://know.dissco.eu/handle/item/164

2. Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1 : https://riojournal.com/article/54280/

3. DPP: https://www.dissco.eu/dissco-prepare/

4. DPP Task 1.1: https://doi.org/10.34960/xhxw-cb79

5. DPP Task 1.2: https://doi.org/10.34960/n3dk-ds60

6. use cases: https://github.com/DiSSCo/user-stories

7. Jacobsen 2020 a, b, c: https://doi.org/10.1162/dint_r_00024

8. Miksa 2019 a, b, c, d : https://doi.org/10.1371/journal.pcbi.1006750

9. Leeflang 2022 a, b, c: https://doi.org/10.3897/biss.6.90987

10. Grieb 2021 a, b, c, d, e, f, g: https://doi.org/10.3897/biss.5.75634

11. CMP T6.4 a, b, c: https://docs.google.com/document/d/1P0QCCtRrd5Cq4JbBWlZMCejEeJ6OUg4-llhstsMGttY/edit?usp=sharing

12. FDOF Public Document Management Register a, b, c: https://docs.google.com/spreadsheets/d/1O1PTLmVWqjQgvJsIcPgaZJe9BntHZt8_QndEG0eRwUY/edit#gid=42354468

13. Islam 2020 a, b, c, d: http://doi.org/10.5334/dsj-2020-050

14. SYNTHESIS projects a, b, c: https://know.dissco.eu/handle/item/175

15. schema.org a, b, c: https://schema.org/

16. repo a, b, c: https://sandbox.dissco.tech/

17. Leeflang 2022 a, b, c, d, e, f, g, h, i, j: https://doi.org/10.5281/zenodo.6832200

18. Hardisty 2022 a, b, c, d, e, f, g, h, i : https://doi.org/10.1162/dint_a_00134

19. Glöecker et al. 2022 a, b, c: https://know.dissco.eu/handle/item/490

20. DD-132 a, b, c, d, e, f, g, h, i, j, k, l: https://naturalis.atlassian.net/browse/DD-132

21. DD-133 a, b, c, d: https://naturalis.atlassian.net/browse/DD-133

22. DD-5 a, b, c, d, e, f, g, k: https://naturalis.atlassian.net/browse/DD-5

23. DD-113 a, b, c, d: https://naturalis.atlassian.net/browse/DD-113

24. DD-13 a, b, c, d: https://naturalis.atlassian.net/browse/DD-13

25. Digital Specimen Repository a, b, c, d, e, f, g, h, i, j: https://docs.google.com/document/d/1XvQBRHeo2Kt-K3j-tVNLAh_RJdFm2oEpR8Q2SjutTzs/edit

26. DD-135 a, b, c, d: https://naturalis.atlassian.net/browse/DD-135

27. DD-26 a, b, c, d, e, f, g, h: https://naturalis.atlassian.net/browse/DD-26

28. DD-18 a, b, c, d: https://naturalis.atlassian.net/browse/DD-18

29. ELViS Code Evaluation a, b, c, d: https://docs.google.com/document/d/165ZdDu1lDdgbR3wgZsqcl8uXm9tylimf295Wh4ugDms/edit

30. DD-16 a, b, c, d, e, f, g, h, i, j, k, l: https://naturalis.atlassian.net/browse/DD-16

31. DD-6 a, b, c, d, e: https://naturalis.atlassian.net/browse/DD-6

32. Latimer Core a, b, c, d, e, f, g, h: https://doi.org/10.3897/biss.6.91159

33. DD-119 a, b, c, d, e, f, g, h: https://naturalis.atlassian.net/browse/DD-119

34. DD-32 a, b, c, d: https://naturalis.atlassian.net/browse/DD-32

35. Collins 2018 a, b, c: https://doi.org/10.2777/54599

36. Hardisty 2021 a, b, c, d : https://doi.org/10.3897/rio.7.e67379

37. Addink 2022 a, b, c, d, e, f:
    https://docs.google.com/document/d/1eD1cPkk0bIzbuezwbVj89DG5AiCMyWHFQk
    hkik9A7M4/edit#heading=h.ghnb1aaw0ji2

38. Broeder 2021 a, b, c: https://doi.org/10.5281/zenodo.4651421

39. CWL a, b, c: https://www.commonwl.org/

40. RO-Crate a, b, c: https://www.researchobject.org/ro-crate

41. EUDAT license tool a, b, c: https://eudat.eu/services/userdoc/license-selector

42. JSON Patch a, b, c: https://datatracker.ietf.org/doc/html/rfc6902/

43. Khan 2019 a, b, c: https://doi.org/10.1093/gigascience/giz095

44. DiSSCo Digitisation Guides a, b, c, d: https://dissco.github.io/

45. Schultes 2022 a, b, c: https://doi.org/10.3389/fdata.2022.883341

46. Wittenburg 2022 a, b, c, d, e: https://doi.org/10.5281/zenodo.5872645

47. MIDS a, b, c: https://github.com/tdwg/mids

48. DD-10 a, b, c, d: https://naturalis.atlassian.net/browse/DD-10

49. trust model a, b, c, d:
    https://docs.google.com/document/d/15tM6qQdP4PPD0_YhQy_0XLq43DX_tE8lpO6
    kscY6ORM/edit

50. agile product roadmap a, b, c, d:
    https://www.atlassian.com/agile/product-management/roadmaps

51. BioDT a, b, c, d: https://biodt.eu/

52. BGE a, b, c, d, e: https://biodiversitygenomics.eu/

53. Webster 2021 a, b, c: https://doi.org/10.3897/biss.5.75736

54. mindat.org a, b, c: http://mindat.org/

55. Le 2014 a, b, c: https://doi.org/10.48550/arXiv.1405.4053

56. Devlin 2019 a, b, c: https://aclanthology.org/N19-1423

57. Bakker et al. 2018:
    https://know.dissco.eu/bitstream/item/79/1/Deliverable%20D2.1_ICEDIG_%20Inven
    tory%20of%20criteria%20for%20prioritization%20of%20digitization%20of%20col
    lections%20focussed%20on%20scientific%20and%20societal%20needs%20%28
    1%29.pdf

58. Hardisty 2019: https://doi.org/10.5281/zenodo.3532937

59. Hardisty 2022: https://doi.org/10.1093/biosci/biac060

60. Lannom 2021: https://drive.google.com/drive/folders/1A4vbovXI0h-x_B66XfI12TzmXaB2Lay1

61. Consultation on the Convergence of Digital and Extended Specimens: https://www.dissco.eu/dissco-consultation-convergence-digital-extended-specimen/

62. BioDT WP5 a, b: https://zenodo.org/record/6841357#.Y9O232CZPIU

63. BICKL WP7: https://doi.org/10.3897/rio.8.e81136

64. MIDS/SYNTHESY+ NA4.1: https://doi.org/10.3897/rio.5.e46404

65. Islam 2022: https://doi.org/10.3897/rio.8.e93816

66. data plus method/workflow preservation, aggregation and sharing: https://www.researchobject.org/ro-crate/

67. Biodiversity Community Integrated Knowledge Library: https://bicikl-project.eu/

68. MIDS: https://github.com/tdwg/mids

69. Deliverable 5.2, Digital Specimen Data Model: https://know.dissco.eu/bitstream/item/401/1/DiSSCo%20Prepare%20Deliverable%205.2%20with%20cover%20%284%29.pdf

70. Deliverable 6.2, Implementation of DS Architecture: https://know.dissco.eu/bitstream/item/512/1/DPP%20D6.2%20Technical%20Architecture%20v_1_2.pdf

71. this report: https://doi.org/10.3897/rio.8.e97374

72. BiCIKL project, 7.3: https://doi.org/10.3897/rio.8.e81136

73. here: https://doi.org/10.3897/rio.6.e57602

74. MOBILISE: https://www.mobilise-action.eu/

75. final deliverable: https://know.dissco.eu/bitstream/item/510/1/WP3_D3.2%20DiSSCo%20Digitisation%20Guides%20Website%20-%20Consolidating%20Knowledge%20on%20Collections%20Mobilisation.pdf

76. dissco.github.io: http://dissco.github.io

77. Catalogue of Life: https://www.catalogueoflife.org/

78. ChecklistBank: https://www.checklistbank.org/

79. DiSSCo Knowledgebase: https://know.dissco.eu/

80. ELViS: https://www.dissco.eu/elvis-is-in-the-dissco/

81. CORBEL Shared Services for Life-Science: https://www.corbel-project.eu

82. DARIAH-EU: https://www.dariah.eu/helpdesk/

83. CLARIN: https://www.clarin.eu/content/support

84. Lifewatch: https://www.lifewatch.eu/help-desk/

85. Islam 2021 a, b: https://doi.org/10.5334/dsj-2020-050

86. MIDS: https://www.tdwg.org/community/cd/mids/

87. INSDC a, b: https://www.insdc.org/

88. GBIF a, b: https://www.gbif.org/

89. Borsch 2020: https://doi.org/10.3897/rio.6.e50675

90. EOL-TraitBank: https://eol.org/traitbank

91. Collins 2018: https://doi.org/10.2777/54599

92. DiSSCo repo: https://sandbox.dissco.tech

93. openDS: https://github.com/DiSSCo/openDS

94. DMF a, b ; https://modelling.dissco.tech

95. Fichtmueller 2021: https://doi.org/10.34960/wn2h-9g16

96. D6.2 Implementation and construction plan of the DiSSCo core architecture: https://know.dissco.eu/bitstream/item/512/1/DPP%20D6.2%20Technical%20Architecture%20v_1_2.pdf

97. syncing basic digital record information, diversity of CMSs, evolving and adapting the CMSs: https://docs.google.com/document/d/1KrNKGWy36oJxoS7jl_cnvVfstp0mK3x1/edit#heading=h.3dy6vkm

98. PID service: https://docs.google.com/document/d/1PPyne6GkWYCb72aYBTNw1TKrWpRy9vK6/edit#heading=h.41mghml

99. a migration document: https://docs.google.com/document/d/1Togkjmmod0LdibEwJhmuyEB11R_wxA-5dXIzXkk0buY/edit?usp=sharing

100.   A call for expressions of interest: https://docs.google.com/document/d/1KygTYz_wyfGQRNJ3UXRFR1iBR7Soh0I_ELrpz77iEG0/edit?usp=sharing

101.   DPP D6.1 a, b: https://doi.org/10.34960/50b9-kj05

102. DPP D6.2 a, b: https://doi.org/10.34960/366d-sf49

103. Data Architecture: https://docs.google.com/document/d/1XzJwbV2zAGpbHE0fl1coi81B5IEV79AU-9-gFbzc7Yl/edit?usp=sharing

104. online consultation: https://discourse.gbif.org/t/converging-digital-specimens-and-extended-specimens-towards-a-global-specification-for-data-integration-phase-1/2394

105. ARISE: https://www.arise-biodiversity.nl/

106. BiCKL D7.1: https://docs.google.com/document/d/1vLxoaDqxCuvEfrzLPrBjHlaTmFsYEG46/edit

107. test Handle system: https://github.com/DiSSCo/handle-server-management

108. Data Management Plan: https://doi.org/10.5281/zenodo.3532937

109. European Open Science Cloud (EOSC) PID Policy: https://doi.org/10.2777/926037

110. research and consultation: https://doi.org/10.3897/rio.7.e67379

111. AARC Blueprint: https://aarc-project.eu/architecture/

112. demo: https://login.dissco.eu/auth/

113. Life Science Login: https://www.eosc-life.eu/news/launch-of-ls-login/

114. EOSC Life: https://www.eosc-life.eu/

115. working on: https://naturalis.atlassian.net/browse/DD-13

116. SYNTHESYS deliverable 6.2: https://docs.google.com/document/d/1DS18n1qOWsLwJqKTk5PbFXgp2iYZg9VY1SCGFa2XpbY/edit?usp=sharing

117. JSON:API: https://jsonapi.org/

118. JSON-LD: https://json-ld.org/

119. Digital Object Interface Protocol: https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf

120. here: https://sandbox.dissco.tech/api/swagger-ui/index.html

121. UCAS: https://sandbox.dissco.tech/

122. proof of concept: https://github.com/DiSSCo/unified-curation-and-annotation-service

123. 2022 conference of the Taxonomic Databases Working Group (TDWG), today's Biodiversity Information Standards: https://doi.org/10.3897/biss.6.90987

124. pilot task: https://github.com/DiSSCo/openrefine-cordra-uploader

125. ELViS: https://elvis.dissco.eu/

126. digitisation on demand: https://naturalhistorymuseum.blog/2021/02/10/digitisation-on-demand-mobilising-natural-history-data-to-address-societys-biggest-challenges/

127. SYNTHESYS+ a, b: https://www.synthesys.info/

128. DINA: https://www.dina-project.net/wiki/Welcome_to_DINA

129. ColHelper: http://colhelper.mnhn.fr/

130. DPP Roundtable 2: CMS DS: https://docs.google.com/document/d/1kPE_x0HVUVA1mE7HLqrPcDib0B3pRbY3kWT0uD9Jp5A/edit#heading=h.bzq9pc5ppwe

131. SYNTHESYS+ deliverable 6.2: https://docs.google.com/document/d/1DS18n1qOWsLwJqKTk5PbFXgp2iYZg9VY1SCGFa2XpbY/edit?usp=sharing

132. JRA1: https://www.synthesys.info/joint-research-activities/jra-1-synthesys-3.html

133. Picturae: https://picturae.com/en/

134. Github: https://github.com/DiSSCo/elvis

135. front end: https://github.com/DiSSCo/elvis-frontend

136. back end: https://github.com/DiSSCo/elvis-backend

137. Kotlin: https://kotlinlang.org/

138. extracting text and data from images ref: https://doi.org/10.1162/dint_a_00134

139. Galaxy workflow platform: https://galaxyproject.org/

140. MVP: https://github.com/DiSSCo/SDR

141. Several challenges: https://github.com/DiSSCo/SDR/issues/59

142. the SDR code: https://naturalis.atlassian.net/browse/DD-6

143. interactive dashboard: https://rebrand.ly/synth-cdd

144. TDWG Latimer Core: https://www.tdwg.org/community/cd/

145. Minimum Information Digital Specimen (MIDS): https://www.tdwg.org/community/cd/mids/

146. current prototype: https://rebrand.ly/synth-cdd

147. Wieczorek 2022: https://docs.google.com/presentation/d/1x-UO5KxcVcUQvSLgKNpjYrQ5q7CZz1Pa1 GXtLONLJb0/edit#slide=id.ge1d059f51d_0_463

148. FAIR Digital Objects Forum: https://fairdo.org

149. DestinE: https://digital-strategy.ec.europa.eu/en/policies/destination-earth

150. Addink 2022: https://docs.google.com/presentation/d/1XWTI_7II7yB9OFqNZiDkPH4BNHiFm-qwx GTO3uHv9sM/edit#slide=id.p1

151. de Smedt 2020: https://doi.org/10.3390/publications8020021

152. global open science cloud: https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/

153. machine learning-based annotation and feature extraction: https://doi.org/10.1162/dint_a_00134

154. Groom 2021: https://doi.org/10.37044/osf.io/93qf4

155. GeoCASe: https://geocase.eu/

156. The iGA approved model 3 as the most suitable model for DiSSCo-ERIC: https://dissco.teamwork.com/app/files/12158965

157. MS7.5: https://know.dissco.eu/handle/item/474

158. Resources: https://www.dissco.eu/resources/

159. Procurement Strategy (MS8.4): https://know.dissco.eu/bitstream/item/504/1/DPP_WP8_MS8.4_Procurement%20Str ategy%20and%20Policy.pdf

160. Stakeholder Analysis (MS8.6): https://know.dissco.eu/bitstream/item/505/1/DPP_WP8_Ms8.6%20_%20Identifying %20indicators%20for%20alignment.pdf

161. Deliverable 4.5: Cost analysis of transcription methods; https://doi.org/10.3897/rio.6.e56211

162. Deliverable 8.1: conceptual design blueprint for the DiSSCo digitisation infrastructure: https://riojournal.com/article/54280/

163. Deliverable 8.2: Cost Book of the digitisation infrastructure:
https://riojournal.com/article/58915/

164. 30 indicators:
https://docs.google.com/document/d/1QlGG-23__fmqEDcT44NUXqK80fA-QC2xw5J
AdF-j8AA/edit#heading=h.9xpvixjf3zqj

165. DiSSCo Prepare WP2 –
Ms2.5: https://know.dissco.eu/bitstream/item/466/1/DPP_WP2_Ms2.5_Landscape%
20Analysis.pdf

166. D2.1 Training strategy report: https://doi.org/10.34960/zm3z-s230

167. MS3.1: https://know.dissco.eu/handle/item/157

168. MS3.3: https://know.dissco.eu/handle/item/159