DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at info@dissco.eu.

Title
DiSSCo Prepare Milestone report MS5.2 "Implementation of concepts for sustainability of services, CMS, and overall TRL"

Author(s)
Mareike Petersen (MfN),
Sabine von Mering (MfN),
Julia Pim Reis (MfN),
Falko Glöckler (MfN),
Claus Weiland (Senckenberg),
Wouter Addink (Naturalis),
Robert Cubey (RBGE),
Anton Güntsch (BGBM),
David Fichtmueller (BGBM),
Mathias Dillen (Meise)

Identifier of the author(s)
MP: https://orcid.org/0000-0001-8666-1931
SvM: https://orcid.org/0000-0003-2982-7792
JPR: https://orcid.org/0000-0002-5357-6148
FG: https://orcid.org/0000-0002-7127-2738
CW: https://orcid.org/0000-0003-0351-6523
WA: https://orcid.org/0000-0002-3090-1761
RC: https://orcid.org/0000-0001-7902-3843
AG: https://orcid.org/0000-0002-4325-4030
DF: https://orcid.org/0000-0002-0829-5849
MD: https://orcid.org/0000-0002-3973-1252

Affiliation
Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

Contributors

Publisher
DiSSCo Prepare

Identifier of the publisher

Resource ID
https://doi.org/10.34960/64bh-7y50

Publication year
2022

Related identifiers

Is it the first time you submit this outcome?
Yes

Creation date
18/01/2022

Version
1

Citation
Petersen M. et al. (2022): DiSSCo Prepare Milestone report MS5.2 "Implementation of concepts for sustainability of services, CMS, and overall TRL". https://doi.org/10.34960/64bh-7y50

Abstract
The Milestone Report 5.2 “Implementation of concepts for sustainability of services, CMS, and overall TRL” is a project outcome of the Work Package 5 “Common Resources and Standards” for the DiSSCo Prepare Project. The report drafts criteria and requirements for the sustainability of digital services developed by the DiSSCo Research Infrastructure (RI) and its community. The demands defined and discussed are

important to ensure long-term accessibility and functionality of the services and thus are also essential for enabling the DiSSCo RI and its core features.

Content keywords
scientific

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP5

Project output
Milestone report

Deliverable/milestone number
MS5.2

Dissemination level
Public

Rights

License
Attribution 4.0 International (CC BY 4.0)

Resource type
Text

Format
PDF

Funding Programme
H2020-INFRADEV-2019-2

Contact email
mareike.petersen@mfn.berlin

H2020-INFRADEV-2019-2
Grant Agreement No 871043

DiSSCo Prepare WP5 – Milestone report

MS5.2 Implementation of concepts for sustainability of services, CMS, and overall TRL

Work package lead: Mareike Petersen (MfN)

Authors: Mareike Petersen (MfN), Sabine von Mering (MfN), Julia Pim Reis (MfN), Falko Glöckler (MfN), Claus Weiland (Senckenberg), Wouter Addink (Naturalis), Robert Cubey (RBGE), Anton Güntsch (BGBM), David Fichtmueller (BGBM), Mathias Dillen (MeiseBG)

Abstract

The Milestone Report 5.2 “Implementation of concepts for sustainability of services, CMS, and overall TRL” is a project outcome of the Work Package 5 “Common Resources and Standards” for the DiSSCo Prepare Project. The report drafts criteria and requirements for the sustainability of digital services developed by the DiSSCo Research Infrastructure (RI) and its community. The demands defined and discussed are important to ensure long-term accessibility and functionality of the services and thus are essential for enabling the DiSSCo RI and its core features.

Keywords

long-term, governance, service level agreements, tools, functionality, DiSSCo RI

INDEX

Introduction & background

The Distributed System of Scientific Collections (DiSSCo) aims to digitally unify all European natural science assets under common access, curation, policies and practices (according to the FAIR principles). To realize this, various services are currently under development in order to create a unique access point for integrated data analysis and interpretation. The access to these services and related information should not be restricted to a short-term period as e.g. the duration of a project but be provided in a sustainable manner. For this report, we define sustainability as long-term, persistent and secured access to the DiSSCo core services and their functions and at the same time the interoperability and re-use (if possible) of services and tools developed and provided by the community. The importance and challenge of the more comprehensive definition of the term sustainability is discussed in the last sections of this report.

Criteria and required information for ensuring sustainability, challenges and approaches implementing these demands as well as bodies of the DiSSCo Research Infrastructure (RI) to be involved in this process are tackled in this report. Since the DiSSCo RI is currently researching, piloting and testing different systems and their implementation, the concepts for sustainability described in this report are preliminary compilations. A more mature concept can be prepared with an increasing developmental stage and higher readiness level of the services. The document will be reviewed and updated periodically if needed.

DiSSCo e-Services

DiSSCo's e-Services are digital services provided by the DiSSCo Research Infrastructure (RI). They are developed based on needs of the scientific community, those of other users of the natural science collections and their development is not least prioritized by the collection holding institutions. Based on discussions from a strategic Workshop in December 2021, DiSSCo aims to provide e-Services to support:

i. Authorisation,
ii. Access (physical and digital),
iii. Capacity building for digital transformation, (crowdsourced) Curation,
iv. Data Publication,
v. Digitisation,
vi. Discovery,
vii. Knowledge, and
viii. Support.

Some of these requirements are already covered by tools currently under development (https://www.dissco.eu/services/). The already available e-Services differ in their developmental stage, the hosting institution, the research project and the Technical Readiness Level (TRL). TRLs are indicators of the maturity level of particular technologies Technical Readiness Levels - DiSSCo for DiSSCo RI and split into 9 subsequent levels[1]:

- TRL 1 – basic principles observed
- TRL 2 – technology concept formulated
- TRL 3 – experimental proof of concept
- TRL 4 – technology validated in lab

[1] Technical Readiness Level, HORIZON 2020 – WORK PROGRAMME 2014-2015, General Annex: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

- TRL 5 – technology validated in relevant environment
- TRL 6 – technology demonstrated in relevant environment
- TRL 7 – system prototype demonstration in operational environment
- TRL 8 – system complete and qualified
- TRL 9 – actual system proven in operational environment

Further e-Services currently under consideration might use tools and pilots planned and developed in different DiSSCo-linked projects and will be available for testing in the upcoming years.

The e-Services are embedded in different levels of the overall DiSSCo architecture (Fig. 1). Some e-Services are directly used by end users (Community Services, e.g. ELViS), others are part of the Digital Object Infrastructure (e.g. Digital Specimen Repository NSIDR) or relevant for data repositories and computing infrastructure. Table 1 summarizes the e-Services currently under development, its scope and further parameters.

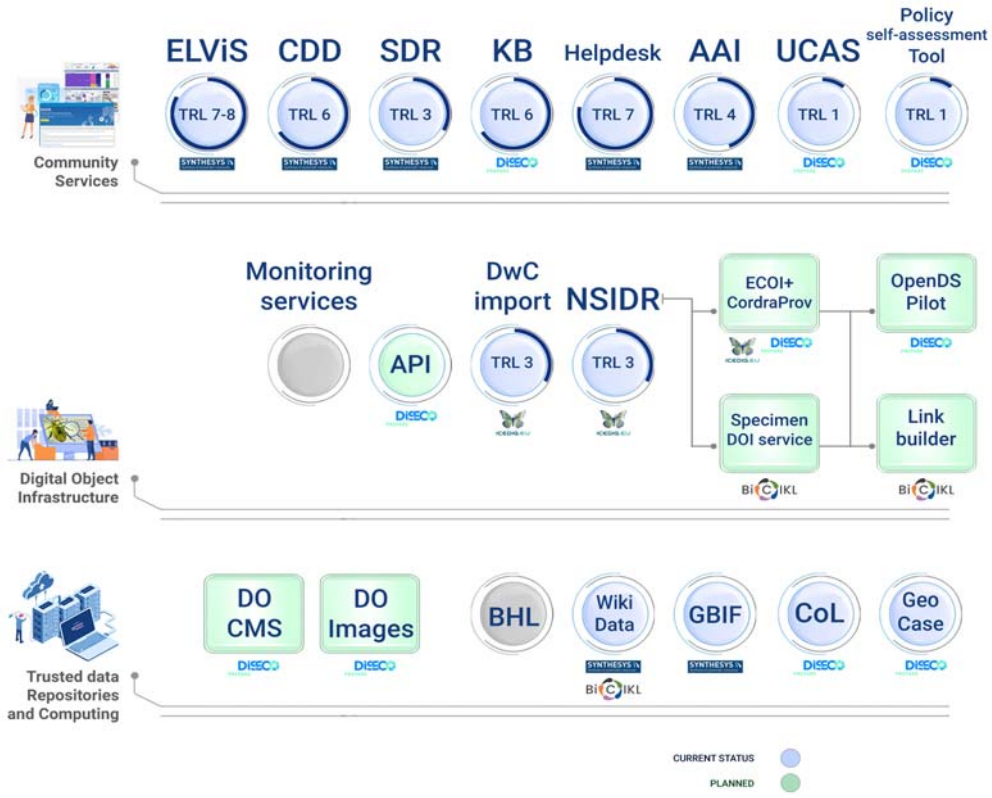

Figure 1: The DiSSCo Architecture. The e-Services can be assigned to one of three levels of the overall DiSSCo Architecture depending on their main scope and application (Figure and describing information at https://www.dissco.eu/services/)

Table 1: DiSSCo e-Services currently under development. The table lists information about the e-Service, the hosting institution, URL (if available), the version and TRL (Technical Readiness Level).

e-Service	Description	Host	URL	Version	TRL
European Loans and Visits System (ELViS)	One-stop shop for access to the collections in Europe, already used for Transnational and Virtual Access in Syn+	Naturalis, Picturae	https://elvis.dissco.eu	Version 1 developed in Synthesys+ Project	7-8
Collection Digitization Dashboard (CDD)	Dashboard visually summarises different collections KPIs across the community through visual elements.	NHM London	https://reprand.ly/synth-cdd	Prototype developed in Synthesys+ Project	6
Specimen Data Refinery (SDR)	Supports industrial scale digitisation including approaches to extract, enhance and annotate data. See also Walton (2020).		NA	Defined & prototyped in Synthesys+ Project	3
DiSSCo Knowledgebase	Central hub to search and browse through documentation related to DiSSCo and any deliverables of DiSSCo-linked projects	MfN Berlin	https://know.dissco.eu	Pilot / Beta Version developed in DPP	7
Helpdesk	Central place for all questions related to DiSSCo e-Services and Access programmes	CETAF, Naturalis	https://jitbit.com/helpdesk	Version 1, used for ElViS in production	7
Authorisation and Authentication Infrastructure (AAI)	Refers to a service and a procedure that enables members of different institutions to access protected information that is distributed on different web servers.	GRNET	https://synthesys.aai-dev.grnet.gr/auth/	Pilot, implemented in Synthesys+	7
Unified Curation and Annotation System (UCAS)	Will provide event-based curation and annotation functions on the Digital Specimen for experts in the community and for machines. Transactions on the data will be stored as well as provenance information related to the curation or annotation events.		-	-	1
Digital Specimen Repository (NSIDR.ORG)	Metadata repository for experimentation with Digital Extended Specimen And DiSSCo-related FAIR digital Objects	Cardiff, Naturalis, Senckenberg	https://nsidr.org/ & https://demo.nsidr.org/	Demonstrator developed in ICEDIG	3
DiSSCo Modelling Framework (DMF)	The DMF will provide modelling capabilities to define data standards used by DiSSCo, such as the Open Digital Specimen (OpenDS). (Internal service)	BGBM Berlin	https://modelling.dissco.tech/	Production Installation, developed in DPP	8

Collection Management Systems across the DiSSCo RI

Collection management systems (CMS) play a key role in the connection between the documentation of metadata on physical specimens deposited in the collection holding institutions and the Digital Specimens curated in or with the help of the DiSSCo RI. CMSs are used to document the inventories of the collections, store data about the specimens, e.g. their collecting events (e.g. gathering site, date, collectors), their scientific identifications and their physical state. In addition, the systems are used to organize transactions within the collections like acquisition, loans and deacquisition. Generally speaking, CMSs and the databases included therein are the electronic collection catalogs. As the digitization of collections proceeds, CMSs also became important by providing a home for the former analog information from collection catalogs, filing cards or books of arrival. There is a huge variety of CMS available, each designed for specific disciplines or purposes, and differing in their functionality, technology and data models. This diversity across different institutions is a challenge for the integration in global infrastructures like the DiSSCo RI (compare Casino et al 2017, Dillen et al 2019).

Implementation of Sustainability

Requirements to ensure sustainability of e-Services

To ensure long-term functionality of developed e-Services and tools, various requirements need to be considered and implemented in any concept for sustainability. Some demands are applicable for all, others are e-Service or tool specific. A list of relevant properties and criteria describing the requirements for sustainability need to be set up. Each e-Service or tool under development can be investigated using this list to estimate its level of sustainability for the DiSSCo RI and the wider community.

Property	**Description**
Purpose and Scope	The purpose and use case of the e-Service or tool needs to be described as concrete as possible, e.g. using a FitSM template[2].
Service Management Plan	A document to concretely describe organisational commitment and responsibility as well as the management process of a service including a Service Level Management (Define, agree, and monitor service levels with customers by establishing service level agreements (SLAs) and supportive operational level agreements (OLAs) [3].

[2] Customized FitSM templete for the DiSSCo RI https://docs.google.com/spreadsheets/d/1prTbjfVkGvIfSOHX3gYrw4vPgqPtEG-2Fi2kJ8_ivCY/edit#gid=495281090

[3] Draft Service Management Plan based on FitSM: https://docs.google.com/document/d/1G-er1Waje35XuF5-yULrMlFf98axKxiqP4YCCpKZ6G8/edit#

Service Level Agreements (SLA)	DiSSCo is a distributed system and many tools and services used by the RI are developed and maintained by partners in different DiSSCo-linked projects or sub-contractors. Service Level Agreements addressing the key aspects need to be negotiated between the RI and the partner. These usually include the technical maintenance of the e-Service but might not be limited to it. There is no blueprint SLA, but they might be based upon the FitSM template[4].
Software Licenses	If possible, the software used for the development of an e-Service should be open source and any products should be available under open licenses (see Albers 2021).
Technical Maintenance	Technical Maintenance such as secured IT infrastructure, automatic backups and security updates needs to be guaranteed. This can be realized e.g. through SLAs and Service Management Proposals.
Technical Readiness Level (TRL)	The TRLs describe the maturity level of a tool or service. A higher level to a certain point in time increases sustainability of an e-Service and contributes to a successful construction of the DiSSCo RI.
Mirroring	Whether an e-Service can be re-used by DiSSCo partners or partners in and outside the Natural Science Collections, its sustainability might be improved by mirroring the systems.
Funding	Each e-Service needs adequate funding for a secured maintenance (see SLA) but also funding for additional developer resources to meet altered requirements. Depending on the e-Service, costs for additional personnel might be required (e.g. curation / quality control in the DiSSCo Knowledgebase, help desk).
Human Capacity	Experienced people (e.g. developer, manager, data scientist, software architect) are required for further development to adopt future user needs and additional or altered requirements (everything in addition to maintenance, compare SLA).
Governance	Each e-Service needs governance. The governing body is in charge to verify this list of requirements, amend where necessary, suggest additional development, or even decide on the retirement of an e-Service. The model of governance, responsibilities and involved parties need to be defined and documented. The governance model of an e-Service is usually described in the Service Management Plan and included in the FitSM template (see above).

Sustainability of Collection Management Systems

In order to identify, prioritize and document connection points and potential dependencies between Collection Management Systems (CMS) and the DiSSCo RI, the work-package "Technical Architecture & Services Provision" (WP6) conducted an Event Storming Workshop[5]. This workshop was organized together with curators, managers, users and developers of CMSs and was intended for a general audience with a technical threshold as low as possible. The goal of the workshop was to brainstorm and aggregate all kinds of events that could occur to a Digital Specimen in both the CMS and the DiSSCo RI (Fig. 2). The aggregated results represent the most important events and help to specify harmonized

[4] SLA template by FitSM: https://confluence.egi.eu/download/attachments/26413993/FitSM_Template_SLA_1.0.docx?version=1&modificationDate=1530098154667&api=v2

[5] Event Storming Method: https://www.eventstorming.com/

API (Application Programming Interface) guidelines and general rules on how to treat these events on both sides to overcome the high effort of connecting the diverse CMSs to the DiSSCo RI. For example, Figure 2 shows the events of (1) creating a new specimen record in a CMS and (2) a change of determinations of a specimen. The events could have easily been identified by looking at the capabilities of the respective CMSs. But by using the Event Storming method it turned out that these events can be triggered by different activities which represent different use cases. So this can be considered in the communication between the CMS and the DiSSCo RI, and thus, in appropriate responses to events, resulting in different types of metadata relevant to the Digital Specimen. Understanding these events and triggers is key to create sustainable recommendations for the API design and to facilitate interoperability.

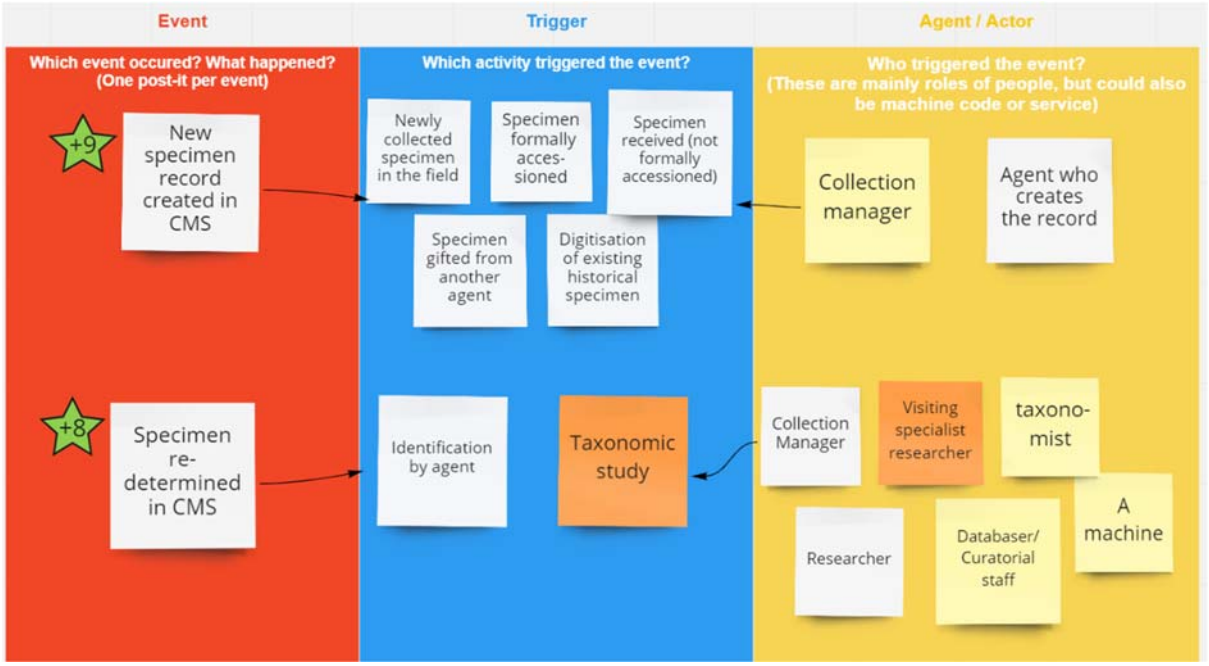

Figure 2 : Screenshot from the aggregated results (partial) from the workshop break-out groups.

The principle of harmonizing different APIs and/or components of existing and new CMSs via common API guidelines has been part of the vision of the DINA Consortium[6] since 2012. This consortium aims to support researchers, collection managers and other users of CMSs with distributed development of an open-source collection management system - the DINA system. As a next step on this journey, the consortium identified that the collaboration between vendors and developers of different open-source CMSs is essential to overcome the obstacles of CMS users. Therefore, as much interoperability and harmonized APIs as possible would be needed. Thus, it should be achieved that components of different systems could be deployed according to the user's needs instead of facing the tradeoff between desired features accommodated in different systems.
An informal meeting (the "Open CMS Collaboration") was organized by the DINA consortium in order to discuss the advantages and potential hurdles of such an endeavor towards a more sustainable way for users to run a CMS in their institutions.

[6] DINA Consortium: https://dina-project.net/wiki/Welcome_to_DINA

The most important aspects for a CMS to be sustainably integrated into an institutional infrastructure can be summarized as follows:

- Cost-benefit: a reasonable cost-benefit ratio for both, the operation and maintenance of the system,
- Sustainability and security of the system: availability and possibility of long-term support for the technical dependencies as well as the specific technical implementation, commitment to active further development by a company or organization,
- Documentation: proper documentation and/or use of commonly used and hence widely supported (ideally open) frameworks and open API specifications like JSON-API and OpenAPI (Swagger),
- Independence of the operating system (e.g. through the use of web based systems,
- Adaptability: flexibility to change along with evolving needs and evolving technology
- User friendliness: a good fit of the system's capabilities with the collection's needs, including enduring staff support, and
- Maintenance of interoperability with the DiSSCo services

Discussion

Challenges and Approaches

The requirements listed above should be considered for the implementation of any concept for sustainability of the DiSSCo e-Services. However, some of the criteria cannot be met since preconditions are not set. Especially in distributed networks such as DiSSCo, different partners or subcontractors lead the development and are in charge of the maintenance of a Service. During the project period the responsibilities and products to be delivered are defined in the Grant Agreement. After a task and / or project is completed, usually SLAs are negotiated among partners to continue the work. In the case of the DiSSCo RI it is obvious to negotiate a SLA between the RI and the Service developing partner or company but this is currently not possible because the formation of a legal entity of the DiSSCo RI is still pending (see also Service Management Proposal[7]).

In the current preparatory phase of DiSSCo, developmental processes of the e-Services are different. Whereas some tasks are foreseen to deliver constructive versions, others only provide pilots. The task planning as well as the final deliverable of a pilot or demo version should comprehend a potential developmental process of how to transform these preliminary products into a constructive version.

Technology is always subject to change and therefore a challenge for big distributed data infrastructures such as DiSSCo. However, various decisions have been taken to minimize these effects. The proposed DiSSCo core data model will be platform and system independent by building upon open standards and in compliance with the FAIR principles[8]. Several of the tools and e-Services use open source software products with open code repositories allowing easy integration of additional requirements to meet new demands.

Whereas the purpose of most e-Services have a low dependency level on their software product (e.g. Digitisation Dashboard, Modelling Framework, Knowledgebase), others like the chosen identifier system strongly depend on a delivering organization. However, those decisions were thoughtfully

[7] Service Management Proposal: https://docs.google.com/document/d/1G-er1Waje35XuF5-yULrMlFf98axKxiqP4YCCpKZ6G8

[8] Fair Principles: https://www.go-fair.org/fair-principles/

considered and the identifier system (DOIs and Handle system) has measurements in place that makes it unlikely that this system is terminated in the next 100 years (Klump et al 2015) and the fact that members of this ISO 26324 Registration Authority (International DOI foundation[9]) are jointly responsible for sustainability of the system. The choice for a FAIR DO[10] infrastructure underpins sustainability because FAIR DO is a technology independent concept. Sustainability of the infrastructure has therefore no dependencies that prohibit future technological changes.

One of the most important bodies driving the development of the DiSSCo RI is the DiSSCo Coordination and Support Office (CSO)[11]. The team can give valuable advice for the different criteria of sustainability and will potentially play a major role in the decision of the future governance of the different e-Services. Another important and active body is the DiSSCo Technical Team[12]. The group consists of domain experts and supervises the distributed development of the e-Services and advises the developing teams if needed. The DiSSCo core e-Services are developed by the team itself or in close collaboration with its members. The DiSSCo Technical Team will be crucial for the establishment of any concepts for sustainability of the technical infrastructure.

The project DiSSCo Prepare has its own Technical Advisory Board (TAB)[13] consisting of independent international experts. They play a key role in identifying specific new challenges and important issues to consider, but might as well give valuable suggestions to the sustainability concepts of DiSSCo e-Services.

Sustainability of information

The term sustainability has been discussed over many decades, and is still addressed today. There are three forms of sustainability: economic sustainability, social sustainability and environmental sustainability which are all correlated (Chowdhury, 2013).

The generic model of the sustainable digital information e-Services illustrates among others how factors such as the information and communications technology infrastructure (comparable to systems and e-Services developed by DiSSCo) affects all forms of sustainability: i. economic sustainability by increasing costs, increasing levels or efforts for ICT equipment for assessing the information; ii. social sustainability by e.g. changes in users' information behavior, culture, social inclusion; iii. environmental sustainability by e.g. increasing emissions due to increasing use of ICT (Chowdhury 2013). Figure 3 summarizes the increasing challenges of information e-Services in terms of all three areas of sustainability.

Within this report, we only tackle sustainability as stability and long-term functionality with reliable access to the developed tools and e-Services. The sustainability in terms of the three spheres discussed in this section are important and need to be taken into consideration in a later developmental stage of the DiSSCo RI. DiSSCo is committed to open science and uses open source software and operating systems for the development of the e-Services where possible. The decision towards free software and IT infrastructures improves the sustainability of any tool and should be pursued where feasible to save limited and natural resources (Albers 2021).

[9] International DOI foundation: https://www.doi.org/
[10] FAIR Digital Object Forum: https://fairdo.org/
[11] DiSSCo Coordination and Support Office: https://www.dissco.eu/about-us/
[12] DiSSCo Technical Team: https://dissco.tech/about-the-technical-team/
[13] DiSSCo Prepare Governace: https://www.dissco.eu/dissco-prepare-governance/

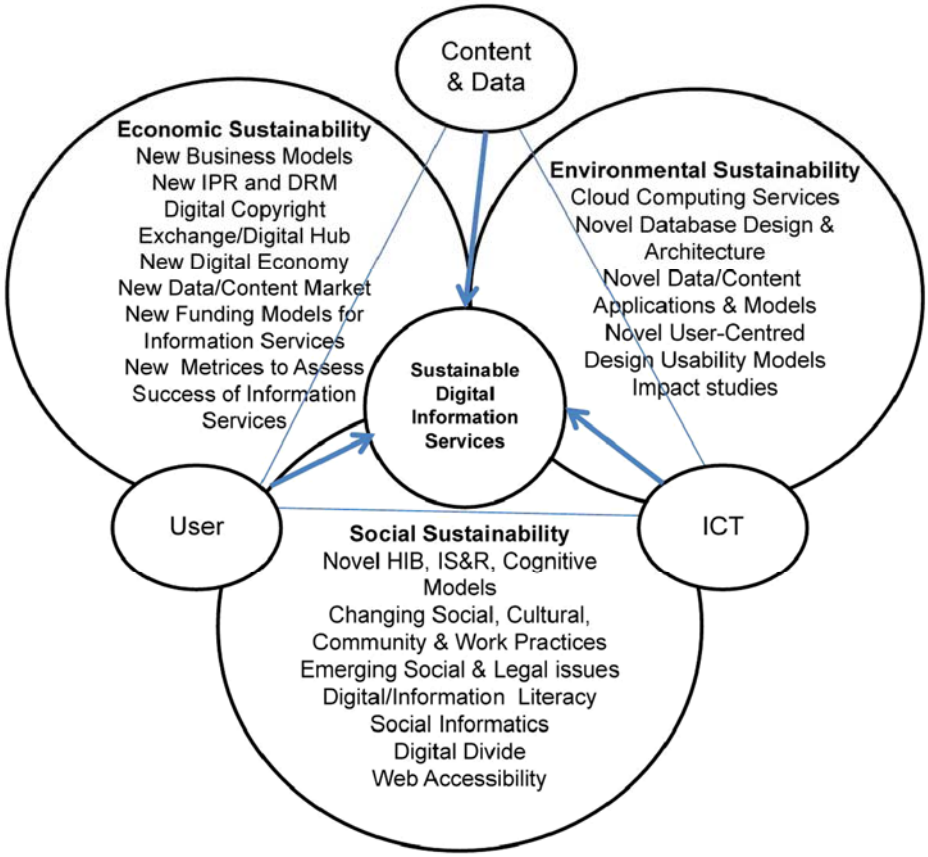

Figure 3: Model of showing emerging research issues and challenges in sustainable digital information services (Fig. 2 in Chowdhury, 2013).

Outlook

This is a first version summarizing the requirements and areas important for implementation of a concept for sustainability for e-Services and tools within the DiSSCo RI. However, e-Services and preconditions are expected to change until the end of this preparatory phase. To retain possible changes in criteria or requirements we will, in addition to this first version, set up a living document (likewise referenced in the DiSSCo Knowledgebase[14]) where any amendments can be added.

References

Albers E (2021): On the Sustainability of Free Software, Free Software Foundation Europe, https://fsfe.org/freesoftware/sustainability/sustainability.en.html (Accessed: 12.01.2022)

Casino A, Gödderz K, Raes N, Addink W,, Koureas D, Hutson A (2019). DiSSCo Partner Capabilities Survey 2017 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.2653708

Chowdhury G (2013): Sustainability of digital information services, Journal of Documentation 69(5): 602-622. https://doi.org/10.1108/JD-08-2012-0104

Dillen M, Groom Q, Hardisty A (2019). Interoperability of Collection Management Systems. Zenodo. https://doi.org/10.5281/zenodo.3361598

[14] DiSSCo Knowledgebase: https://know.dissco.eu/

Klump J, Huber R, Diepenbroek M (2015). DOI for Geoscience Data - How early practices shape present perceptions. Earth Science Informatics. 9. https://doi.org/10.1007/s12145-015-0231-5

Walton S, Livermore L, Bánki O, Cubey RWN, Drinkwater R, Englund M, Goble C, Groom Q, Kermorvant C, Rey I, Santos CM, Scott B, Williams AR, Wu Z (2020) Landscape Analysis for the Specimen Data Refinery. Research Ideas and Outcomes 6: e57602. https://doi.org/10.3897/rio.6.e57602