



ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

**Agreement Number: 777483 / Acronym: ICEDIG
Call: H2020-INFRADEV-2017-1 / Type of Action: RIA
Start Date: 01 Jan 2018 / Duration: 27 months**

REFERENCES:

DELIVERABLE 9.4

Work package WP9 / Lead: NHM

Deadline 31/03/2020

DOI: 10.5281/zenodo.3733033

Deliverable 9.4 Positioning DiSSCo among other research infrastructures

**Hilary Goodson (NHM)
Karsten Gödderz (CETAF)
Ana Casino (CETAF)
Dimitris Koureas (Naturalis)
Stephanie Walton (NHM)**

Executive Summary	2
Introduction	2
DiSSCo scientific audiences and value	3
Impact on Evolutionary Biology, Ecology: Interpreting the tree of life	4
Impact on associated and applied sciences	4
Impact on innovation	5
Impact on Humanities	5
DiSSCo unique value proposition	5
Hard evidence for the natural world	5
Mapping the Landscape	6
DiSSCo Positioning in the wider RI Landscape	6
Approach to DiSSCo positioning	6
Scientific facet	6
Strategic facet	7
Operational facet	7
Managing international complexity in assessing the landscape	7
DiSSCo Positioning within the immediate context	8
Positioning within ESFRI	8
Positioning with CETAF	9
Positioning with GBIF	10
Positioning within EOSC	10
Database and Dashboard Prototype	11
Methodology	11
b) RI elements	11
Dashboard Overview	12
Projects and Collaborations	12
RIs and Services	13
Host Regions	15
Analysis	16
Annex 1	18
Environmental RIs in the ESFRI	18
Annex 2	19
GBIF Statement at the inaugural DiSSCo General Assembly	19



Executive Summary

ICEDIG and DiSSCo need to position their role among other related European and global research infrastructures (such as LifeWatch, GBIF, GEO BON, CoL, EoL, BHL, etc.), build on the services of already existing and planned infrastructures (such as EUDAT, EGI, etc.), and closely follow the ESFRI process. This report contains a detailed analysis of where DiSSCo fits in the global landscape of research infrastructures, specifically in the environmental domain, including the various ways how the already existing RIs can profit from DiSSCo and vice-versa. Possible collaboration paths with existing e-infrastructures and initiatives will form part of the analysis. Possible collaboration paths with existing e-infrastructures and initiatives will form part of the analysis output of T9.5.

Keywords: DiSSCo, Research Infrastructures, Mapping, Partnerships, Global Alliance, integration

Introduction

The ESFRI Environmental Sciences Thematic Working Group report highlighted already in 2010 the need for “an integrating body for the scientific collections” at a European level. However, until DiSSCo appears, the taxonomic backbone, providing the foundational layer at the scale and precision required for further research work was a missing building block and a major limitation to biodiversity and ecosystem knowledge and therefore for the European landscape of environmental and other closely related (as bio-medical) Research Infrastructures (RIs). All of them need to be equipped with adequate data. This is also true for adjacent user communities e.g. in the cultural domain and beyond where field books, archives, illustrations etc. constitute assets that they address. DiSSCo will allow Europe’s researchers and technology professionals to share and reuse the data linked to collections across disciplines and borders. It will mobilise and harmonise science collection data (collection metadata, traits, images, metabolites, nucleotide sequences, distribution or ecological information) and make them available as part of a highly connected linked-data graph.

This report seeks to develop a framework for identifying parallel or related research activities to contextualize and further evaluate the proposing, refining, agreeing and implementing of shared projects and project components towards an aggregated convergence also at international level, in alignment with sister initiatives to DiSSCo. These continental groups are at various stages of maturity in delivering digital collections networking, infrastructure and research activities closely connected to the objectives of DiSSCo. This report seeks to identify all those related pillars of activity and map out their complementarity and define at national



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



level and above, a framework of analysis to be drawn upon as part of DiSSCo's implementation plan.

Such a mechanism would benefit from further development and maintenance to increase opportunities for collaboration and sustainable development of tools and infrastructure. A key priority would be to establish an effective model for supporting such project planning and implementation. With the ICEDIG project coming to an end as of March 2020, these tasks will be picked up by [DiSSCo Prepare](#), the DiSSCo Preparatory Phase Project, specifically in the tasks T8.3 “DiSSCo stakeholder engagement” and T8.4 “Dissemination, Outreach and Advocacy”. Additionally, the DiSSCo General Assembly established a high-level task group to support these activities.

Research Infrastructures across the environmental domain, and specifically in relation to bio- and geo-diversity, are already partnering together on a range of projects that aim to expand cooperation and deliver solutions that can be adopted by all partners. These projects interact through investments, organisations, activities, and tools, that can be envisioned as a multi-dimensional matrix. Collectively these address a series of needs from technical infrastructure to policy recommendations, that all together contribute to ensure the imperative for end to end data mobilisation.

This analysis provides a contextual framework to aid in a more comprehensive assessment of necessary partnerships as we take the ICEDIG blueprint forward, and ensure that we understand the evolving technical and policy landscape that the DiSSCo research infrastructure is positioned in. This will capitalise on efficiencies and investment where there are synergies and help raise awareness of existing components that would enable us to most efficiently build the DiSSCo RI. Such a perspective will be complemented by taking into account a number of international initiatives and organisations complementary to DiSSCo which are developing in regions outside of the EU.

All those other actors, at European level and internationally, include similar digitisation initiatives of national collections, but it is also necessary to consider those projects and research infrastructures that, while not directly involved in biodiversity, run alongside it and, through digitisation, offer greater opportunities for collaboration and coordination. This is aligned with Task 7.3 and the need to develop a common digital research agenda, not just within bioinformatics and digitisation, but across wider sectors.

DiSSCo scientific audiences and value

DiSSCo as a pan-European Distributed System of Scientific Collections looks into the digital re-unification of all natural science collection assets and their transformation to an integrated knowledge base, along with the expertise and policies necessary to maximise impact to science and society. All collections provide base-line data in time and space to reconstruct the history of life on Earth, to monitor changes and model it. DiSSCo aims to translate this



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



enormous set of information into digital components of a larger, interconnected biodiversity fabric.

Impact on taxonomy and systematics

Collection objects have data that document the specimens, including taxonomic name, geographic location, field notes, logbooks and illustrations. Technological innovations have furnished this documented information with additional components, from frozen tissue samples and DNA barcodes to whole genomes, proteomes and metabolomes. Advances in informatics have also complemented that information with 2D and 3D pictures. Globally, labels, images, traits, spatial, morphological, taxonomic, molecular and chemical data constitute the information associated with the physical specimen worth exploiting.

DiSSCo will allow distributed collaboration, using common tools and standards to construct, curate and assess a tree of life based on the fullest available evidence. Knowledge of the diversity of species and their distribution in the biosphere is foundational to biology, ecology and conservation. However, although taxonomic research has made enormous progress in naming and describing species, current progress is insufficient and remains scattered. Collections contain a large proportion of specimens still remaining to be described, but these are distributed across hundreds of institutions and relevant expertise is also highly localised. Digital access to rich data on all specimens, and efficient tools for exploring, organising and curating these digital representations, are required to give experts direct and reliable access to taxon-based information they need. The impact on the acceleration of species descriptions will be visible almost immediately after DiSSCo starts operating, though the closure of current huge knowledge gaps remains a long-term goal.

Impact on Evolutionary Biology, Ecology: Interpreting the tree of life

A robust tree of life will facilitate prediction of the biological requirements and characteristics of each species and of their expected function within ecosystems. It will open fresh areas for research into the forces driving evolution and speciation. DiSSCo will serve not only as the framework for interpreting and validating species data from other sources, but will also deliver the diagnostic information required for novel approaches and technologies for accelerated field identification of species, and contribute to the development of datasets at sufficient scale to support regular monitoring, trend analysis and future prediction. Direct knowledge of species biology will be complemented with phylogenetically derived traits to enable modelling of ecosystem function at diverse scales.

Impact on associated and applied sciences

For all research associated with bio- and geodiversity, like life sciences, biomedical, pharmaceutical and agricultural research, environmental sciences and material science, the data and services provided by DiSSCo hold great potential. Cross-domain collaborations and context-aware interactive services combining collection data with other data will address



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



complex scientific challenges associated for example with climate change, the spread of pests and diseases or the genetics of complex traits. DiSSCo will provide coordinated data services to tailor them to scientific use cases and priority areas and thus form the foundation for addressing these challenges.

Impact on innovation

DiSSCo will directly support innovation as a result of these detected opportunities for collaboration. It will enable industrial partners to develop new technologies, and science to provide answers to societal challenges faster and more accurately and to jointly find novel pathways for solutions. New products and services are expected from these efforts (mid-term) to foster an intelligent and sustainable biodiversity friendly economy in Europe (long-term). For example, new pharmaceuticals as a result of combining collection data with metabolomic information are to be expected as well as new plant cultivars or animal breeds that meet new requirements or new materials using nature as a model.

Impact on Humanities

Natural science collections are part of the Cultural Heritage of our society. The institutions housing scientific collections also keep archives, extensive libraries, and provenance data. DiSSCo will integrate the historical and contextual metadata of each digitized object - how it was collected, who it was collected and curated by, where it was collected, and how each object fits within broader patterns, practices and sites of collection – into the overall endeavour. Thus, DiSSCo will be a truly interdisciplinary platform where researchers interested in the production or the consequences of the investigation of the natural world can find a cross referenced resource.

DiSSCo unique value proposition

Hard evidence for the natural world

Natural Science Collections have as part of their core business the preservation of the physical assets they hold. These assets (biological and geological collection objects) continue to be the focus of scientific studies. Researchers have been working on extracting information and using it to describe our planet's bio- and geo-diversity. This information is scattered across several scientific resources in both analogue or digital form. From laboratory notebooks to large-scale digital data aggregators.

Unlike human observations in nature, where physical evidence is rarely kept, collections persist through centuries, enabling scientists to retrieve new classes of information later in time. Collections provide a unique ability to researchers to validate derived specimen observations and measurements, significantly improving reproducibility of scientific observations. A typical example is the introduction of molecular sequencing technologies. Such technologies, in combination with the accessibility to collection objects, enable scientists to retrieve DNA from spatiotemporally unique objects held in collections globally. Future technologies and new insights into science might further allow us to retrieve ever bigger



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



amounts of unique data that will contribute to reconstruct the distribution and evolution of life on this planet.

Mapping the Landscape

The landscape of initiatives that surround DiSSCo at one level or another in the proximity or in a very lateral way is very complex as it gradually is getting crowded and the initiatives mature and become operational, larger and stronger. We have approached this by creating a centralised database in which related infrastructures, projects, initiatives, platforms and portals can be aggregated, compared and assessed. Some special efforts have been made to catalogue the wide number of research infrastructures and projects in Europe aiming at creating an accessible and query-able form of discovery of digital collections that might be relevant to different ways of research. Though some other international initiatives might be identified, the deep insight has been limited to the EU realm..

However, building a fully functional and query-able database is far beyond the scope of this deliverable, and therefore it circumscribes to the first stages of identification and allocation of projects, this report presents a high-level exploration analysis of research infrastructures and projects related to DiSSCo around the world and some introductory references to n the types of platforms and services that are currently available and/or being developed internationally (see page 9 ff.).

DiSSCo Positioning in the wider RI Landscape

Approach to DiSSCo positioning

To position the DiSSCo RI in the wider landscape in which it operates it is essential that we first understand the facets through which interactions can develop between landscape actors. As a research infrastructure, DiSSCo can navigate the landscape through three distinct, yet interlinked facets. The scientific, the strategic and the operational facet.

Scientific facet

Research Infrastructures are set-up to enable and facilitate science across a multitude of users. The vision and mission of those infrastructures is focused on the achievement of aspirational scientific goals. As such the majority of these infrastructures issue regularly detailed scientific programmes, which subsequently are supported through annual operational plans. As scientific aspirations and societal needs grow, researchers face a growing need for access to multiple infrastructures, in order to support their ever more complex data lifecycles and interconnections with data from approaching fields. In this context, there is a significant opportunity for interaction and convergence of infrastructures both at the level of their scientific programmes and at science operation. Developing partly complementary scientific



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



programmes and connected plans can boost investments that target cross-infrastructure interoperability.

Strategic facet

As the landscape of RIs grows with new additions, it is essential that we develop the tools through which we can better describe it and enable users to navigate it at any moment. From a strategic planning point of view, RIs depending on their development phase, can reflect on how they position themselves in the national and international agendas for research and innovation to leverage the required resources for their continuing and sustainable operation.

Operational facet

At operational level, RIs can benefit from the development and use of common resources and means. Investments in reaching, through RI-to-RI collaborations, economies of scope and scale; common procurement efforts; co-development of shared tools and shared engagement and communication activities could be among the areas of operational collaboration between RIs.

Managing international complexity in assessing the landscape

Trying to find and document adjacencies to DiSSCo requires equally identifying equivocations within a complex landscape that accounts for differing political strategies, technological priorities and structures. Within the diversity of related infrastructures, there is high variability in their level of connectedness. We have focused first on mapping the elements within different components of the DiSSCo RI. We looked at domain-specific infrastructures at national or international level as there are a multitude of biodiversity informatics projects, datasets, databases and initiatives at the global level, and many more at national and regional level. We then examined the relevant thematic clusters, including those amongst the environmental research infrastructures.

Following the first attempt made under the ENVRIplus (Environmental Research Infrastructures) project, we have now attempted to further analyse the landscape through a framework centered on DiSSCo's different components and how other RIs may contribute to collectively make up an end-to-end RI. The analysis acknowledges both horizontal and vertical definitions of "research infrastructure" and describes both in order to understand the landscape in which DiSSCo is positioned. Research infrastructure means the facilities, resources and related services that are used by the scientific community to conduct top-level research such as collections and archives, structures for scientific information, enabling information and communications technologies, such as grids, computing, software and communications or any other entity essential to achieving excellence in research.

This can be understood as an interlinking matrix driven from two points. The first driving point is e-Infrastructure which addresses the needs of European researchers for digital services like networking computing and digital data management. This encompasses elements extracted from the European Open Science Cloud, organisation of networking, data mobilisation and



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



technical interoperability. The second driving point is research infrastructure working to advance impact in fields of research through coordinated and strategy-led approaches to convergence as well as policy making within the environmental research domains.

The matrix will consider specifically EU infrastructures though it is being built on top of the report “International landscape analysis of related international research infrastructures supporting collections digitisation” made under WP7 (Milestone MS46 met in December 2019) that presents a high level drawing of infrastructures potentially related and/or with identified connectivity in regards to collections digitisation.

Approach to Aggregation of baseline list

As part of the delivery of MS46, a baseline list of international research infrastructures that support collections digitisation was aggregated from existing internal knowledge. The list was not intended to be comprehensive but to serve as a starting point for researching into the types of services offered, where further data might be found on other infrastructures and how the data might be structured.

This list contained approximately 50 infrastructures that included biodiversity collections as well as adjacent fields - microbiology, chemistry, marine biology, etc. These infrastructures offer services ranging from digitised collections portals to field sites networks and consulting services and vary in size from highly specialised and local to comprehensive and global. Geographic coverage of this initial list included Europe, the USA, China, Mexico and Brazil.

DiSSCo Positioning within the immediate context

There is a further and more direct set of stakeholder relationships that DiSSCo already engages with.

CETAF and GBIF each need to be considered as both have been and do remain critical to the success in further exploiting the global biodiversity mission. Clear remits of these bodies need to be marked out in order to augment each other and ensure we mitigate risk of inadvertent conflict or duplication of effort. Work has already been established to address these points.

The other set of formal stakeholder relationships is with the other ESFRIs in the biodiversity domain, where we will need to develop cross-cutting functionality and also ensure we can collectively voice our domain needs in the development of further Commission-funded infrastructure development.

Positioning within ESFRI

In the ESFRI 2018 Roadmap the field of environmental sciences is divided into four subdomains, each one dealing with a part of the Earth system: atmosphere, hydrosphere (including the marine and oceanic facilities), biosphere (ecosystems) and the geosphere (solid earth). This classification is used throughout this report considering that in Europe, as well as



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



elsewhere, some RIs deal with multiple subdomains and shall not be restricted to just one, as it is the case of DiSSCo.

The number of Environmental RIs in the ESFRI list (*Annex 1*) is relatively high due to the number of disciplines involved. The research areas in environmental subdomains are closely interlinked and several share similar features. Many of the RIs are observational, with distributed facilities located in a wide geographical area and some support experimental facilities. International collaboration is the de-facto mode for many of the Research Infrastructures in this area, as the challenges they address are often global in nature. Similarly, multidisciplinary is often critical to answer environmental challenges.

Positioning with CETAF

CETAF has been central to the design and development of a new overarching RI for natural science collections across Europe, DiSSCo. CETAF plays a pivotal role in the mobilization of community engagement for the DiSSCo initiative and underpins the business case for DiSSCo, bringing more than 20 years of experience to the task of constructing the DiSSCo RI. CETAF will continue to play an essential role in DiSSCo as part of mutual efforts to transform a dispersed and fragmented model for providing access to collections and associated data into an integrated data-driven pan-European research infrastructure.

CETAF has been a leading supporter for DiSSCo through the participation of the CETAF Executive Director in the DiSSCo Coordination Team, and by the CETAF Chair's position on the DiSSCo Steering Committee. This has enabled CETAF to build consensus throughout its membership on DiSSCo's overarching scientific, technical and governance vision. CETAF has sought to align the goals of DiSSCo with the strategic position and objectives of the CETAF community. Moreover, CETAF has worked intensively to inform, disseminate and reach out to stakeholders, acting as a leading advocate for the DiSSCo RI. The long history of collaborative projects, initiatives and outcomes undertaken within CETAF led to the classification of our consortium as a "super-advanced community" by the European Commission. CETAF's participation in the landscape of environmental RIs, like ENVRIplus and the Board of European Environmental RIs (BEERi), has also helped strategically integrate DiSSCo into the biodiversity-geodiversity research landscape.

CETAF will be central to the next phases of the DiSSCo development plan as it strives for the efficient, cohesive and coherent implementation of DiSSCo. For example, by providing the scientific community with relevant and specialised services and advocating for the DiSSCo RI as well as contributing to outreach activities, the coordination of CETAF members' data and the alignment of goals and priorities across the community.

CETAF's long-standing record of collaboration and cooperation will support the development of the DiSSCo RI, ensuring both equitable and sustainable engagement of the community. As a legal entity and politically neutral organisation, CETAF's standing serves the benefit of DiSSCo's governance model, facilitating the smooth transition between the preparation and later phases. CETAF will contribute to the coordination of the national nodes and act as a



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



representative of institutions/consortia that are too small, or otherwise unable to become full members of the DiSSCo RI, thus playing a key role in ensuring full integration of the community and strong commitment (CETAF leads the WP8 for, among others, DiSSCo national nodes engagement in the DiSSCo Prepare project). CETAF's portfolio of MoU's can also be leveraged to ensure the widest possible participation in DiSSCo, across and beyond Europe.

CETAF is in a unique and fundamental position to complement DiSSCo, and vice versa. The CETAF and DiSSCo digitisation agendas are interlinked, with DiSSCo addressing the infrastructure ambitions of CETAF, that are beyond the resourcing and technical capacity of individual CETAF members. DiSSCo bridges this critical gap, supporting open access to natural science collections from across Europe, as well as bringing a big-data perspective to their scientific exploitation in previously unimaginable ways. The implementation of DiSSCo for 2026 provides an exciting new opportunity for the scientific community to exploit the new data generated from natural science collections that will be mobilized through the DiSSCo infrastructure, for taxonomic and systematics research, and beyond, giving the scientific community the potential to address some of the key societal issues of our time (invasive and alien species, food security, emerging diseases, pollution monitoring, responses to climate change).

Positioning with GBIF

GBIF is a mature governmental-based infrastructure with global membership, developing training and outreach programs and aggregating, managing and serving out complex biodiversity data. Through close collaboration with GBIF and other global partners, DiSSCo can take advantage of GBIF's experiences and existing infrastructure, as appropriate, and continue to work jointly under the banner of the *alliance for biodiversity knowledge* during the DiSSCo Prepare phase and beyond. DiSSCo joined the GBIF network as an associate participant in 2019. On the other hand, GBIF recently participated as a guest in the inaugural General Assembly of DiSSCo and contributed with an intervention (full statement in Annex 2) in which a shared goal was highlighted: "to provide FAIR data for research and policy decision makers". Equally, close collaboration was promoted to develop "critical biodiversity knowledge" and special relevance was given to the collaboration between national nodes from DiSSCo and GBIF to strengthen alignment, joint developments for data mobilization, leveraging of resources, and new data providers engagement.

Positioning within EOSC

A clear and concise communication of the concept of EOSC is required since DiSSCo services will be integrated into this European platform for data services provision (and data).

While DiSSCo connects to EOSC at a very primary level through the ENVRI-FAIR clustering project, specific services from DiSSCo to be included in the EOSC Catalogue are only at their conceptual stage and need further development. DiSSCo has just started its preparatory phase and needs to become operational before the envisaged services for the research community and beyond can be delivered.



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



In the meanwhile, DiSSCo needs to ensure adequate linkage to the EOSC virtual working environment for scientific research, to its long-term open data archiving facilities with high performance storage and computing services as well as to other services that will be at the basis of the so-called Minimum Viable Ecosystem that - if it accommodates the needs of DiSSCo - could be adopted by the DiSSCo RI. Conversely, DiSSCo will highlight the characteristics that underpin its services provision and will ensure translating those to the foreseen EOSC business and governance model for participating infrastructures. DiSSCo will reinforce the idea of a feasible and successful EOSC based on a research community-centred bottom-up approach to which DiSSCo is committed.

Database and Dashboard Prototype

We have opted for a dashboard to present the results in a visual, easy-to-understand, and appealing manner that offers a high-level representation of the huge set of information contained in the database. By making it query-able it can respond to multiple user needs and purposes. It can then easily be transferred to other types of reports and support informed decisions on how to address connection, joint efforts and shared actions in relation to existing RIs. The data visualised in this dashboard is a *representative sample* of 35 research infrastructures and platforms that bear some relation to DiSSCo. While there are many more to be catalogued and incorporated, this sample set provides the opportunity to 1) explore the types of data that can be collected and visualise, 2) the insights that can be revealed by visualising the data and 3) the ways that a future user could search for and identify RIs that offer the potential for collaboration and synergy.

Methodology

a) RI selection

Taking the analysis of the international landscape made with the report for MS46, we proceeded to select 35 RIs considering the field they operate in in relation to DiSSCo. Of course, only those RI that in one way or another collaborate among each other are taken into consideration. Others that might work rather in isolation were discarded from the original list reported in MS46.

Online meetings were held to identify their major and most relevant features and equally the services they offer to the research community. The information was completed with internet searches, direct conversations with peers and other (physical) meetings that took place under other related ICEDIG tasks, allowing us to talk to relevant related infrastructures.

b) RI elements

One of the chief difficulties in creating a database of research infrastructures is the wide variety of terminology, interactions, projects, platforms and relationships that exist in the ecosystem.



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Categorising every entity along consistent definitions can be difficult in some cases where definitions don't apply. However, this categorisation model represents an early attempt to draw equivalencies across research infrastructures.

- **Research Infrastructure:** A collection of institutions that provide resources and services for research communities to conduct research; comprised of multiple coordinating institutions that collectively contribute to the development and delivery of projects, resources and services.
- **Project:** Funded multi-year initiatives worked on by collections of research infrastructures and institutions to accomplish an established set of objectives including tools/services development, networking and integration, research, operations and governance of an RI, etc.
- **Service Category:** A collection of common services offered by research infrastructures to their community such as data analysis services, data collections services or research services.
- **Service:** A specific service offered by a research infrastructure to their community that sits within a service category such as a data portal, field research sites, web APIs and taxonomic name recognition.
- **Host Region:** The country or region that is responsible for managing and executing a research institution. This is separate from member countries or funding bodies that *fund* a RI. A host region is where the people and institutions are based that see to the day-to-day management of the RI.

These same definitions were later used in the structuring process of the dashboard.

Dashboard Overview

This early dashboard prototype was built with a sample dataset (of 35 RIs) to ensure that the data model would be useful in providing the necessary insights of the Work Package objective. The dashboard is available for viewing and testing on the [ICEDIG website](#) as well as on the [DiSSCo website](#), where it will be kept in the long term.

Projects and Collaborations

As one of the main objectives of the dashboard is to highlight collaborations and synergies between RIs, the opening dashboard page illustrates on which projects RIs are working and when they collaborate with other RIs. *Image 1* shows the first page of the dashboard which tells the number of projects in the database, the number of projects currently in operation and the number for which funding is scheduled to end in 2020.

It also shows the categorisation of projects based on their purpose and objectives. As can be seen in the table, a majority of projects are focused on the development of new services for



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



researchers in their community such as data portals or field site development. Hovering over a project name will show the research infrastructures that are working on the project and a description of the project drawn from the project’s website.

Because this dashboard is only based on a sample set of 35 research infrastructures, the number of collaborations among these 35 is somewhat limited. However, the chord chart on the right illustrates the types of insights that will be available with a more expansive dataset. It becomes clear, for example, that GBIF collaborates with a number of other research infrastructures on various projects and is collaborating with GEO on multiple projects. This chart only shows research infrastructures that are collaborating and excludes infrastructures operating in isolation (according to the limited dataset).

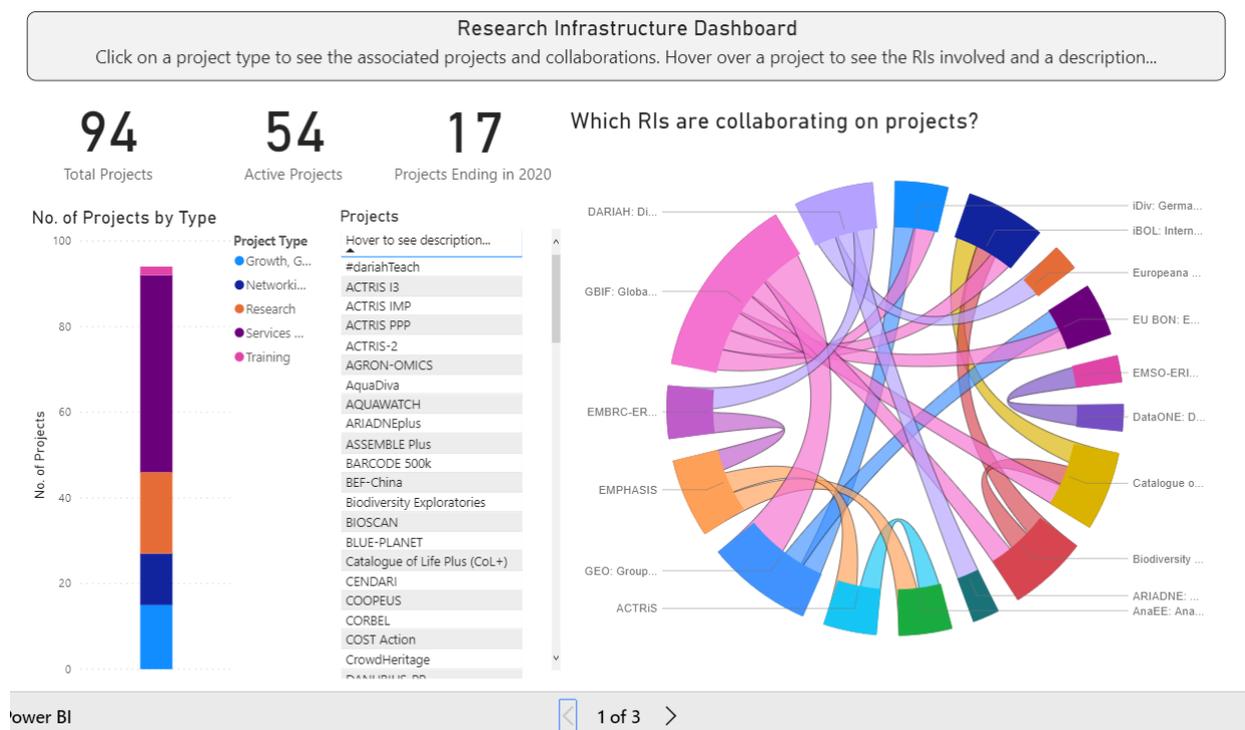


Image 1: Page 1 of the [dashboard](#) illustrating projects and existing RI collaborations.

RIs and Services

Page 2 of the dashboard (*Image 2*) illustrates the wide variety of services offered by the various RIs and the most common services. The bubble chart shows that the two most common services are data portals such as those offered by GBIF, the Atlas for Living Australia and NCBI. These data portals are often accompanied by web APIs and other web services that allow direct interaction with the portal’s data. Field stations are another common service as well as training and consulting.

This dashboard is also interactive so that it is possible to filter the list of RIs based on the field of research. So, for example, if a researcher is looking for an RI focused on a specific subject matter such as marine biology, they could apply the filter and see all the RIs operating in that



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Image 3: The dashboard filtered to show only RIs focused on marine biology and the services they offer (see [dashboard](#) on ICEDIG for more detailed imagery).

Host Regions

Page 3 of the dashboard (*image 4*) shows where different RIs are hosted. Hovering over a country reveals the RIs hosted by that region and clicking on an RI will zoom into the host region.

In the next iteration of this dashboard, it would be beneficial to explore a method for visualising the geography of data coverage for each RI. For example, if a researcher is in need of specific amphibian taxonomic data for Brazil, they could use a map to identify which RIs offer amphibian data for Brazil. A majority of RIs in the sample data included here, technically offer global data coverage such as occurrence data spanning the globe. However, just because an RI technically offers global data coverage does not mean they actually have data on a specific area. Accurately representing this information may require more consideration on whether a visualisation can take this into account and how the data can be aggregated so show the actual available data coverage of each platform.

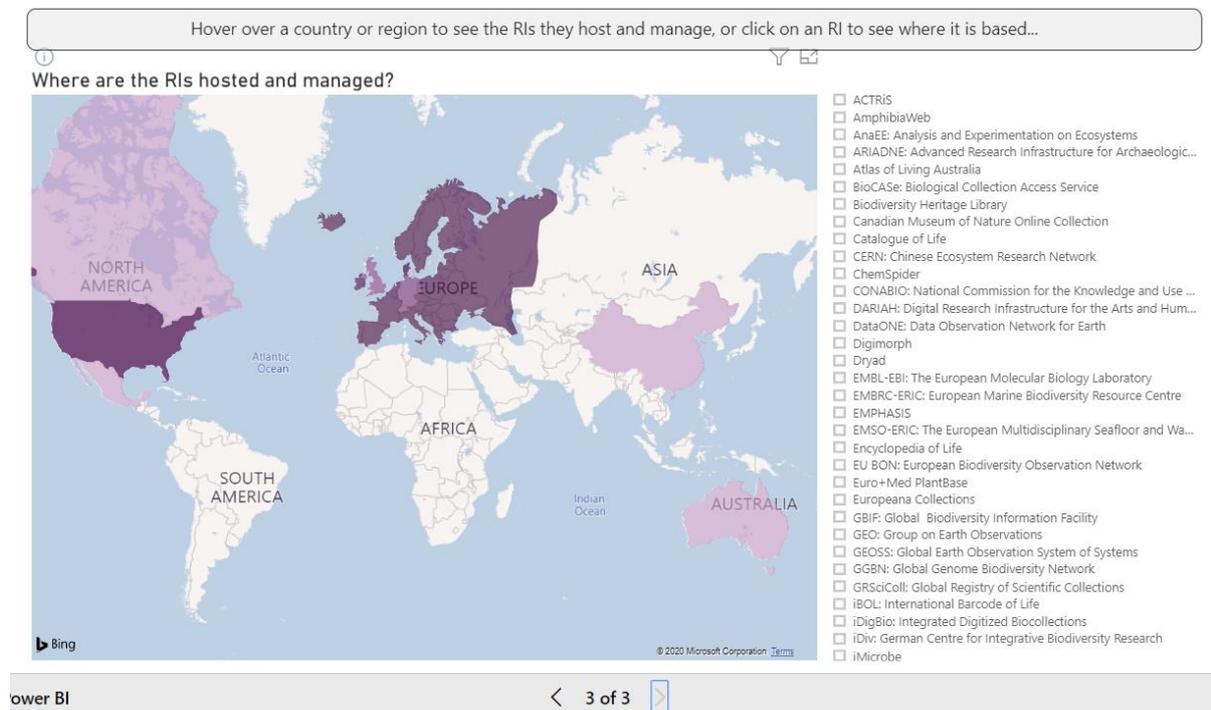


Image 4: Page 3 of the [dashboard](#) illustrating the host regions of each RI.



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Analysis

The primary goal of this task was to create a structure to explore and evaluate possible models for a transparent and trusted collaboration mechanism across Europe—one capable of enabling this community to coordinate its efforts and investments to deliver a shared, open and interoperable knowledge base encompassing all sources of information related to the large biodiversity infrastructure programmes. We recognized the complexity of the existing landscape of institutions, organizations and projects that participate in management and use of biodiversity information. The number of initiatives involved and the overlaps between their various missions and work programmes make it difficult to identify all stakeholders, but this exercise has resulted in a structured model that could be maintained, augmented, and developed over time. The mapping of roles and relationships allows for further layering to be able to extend (geographical and domain) scope, map dependencies, potential reuse of tools and define priorities.

This should serve as a starting point to inform discussions on partnership brokering. Particularly, this should be a reference when examining resource allocation and funding questions concerning services and components that other parts of the community recognize as critical elements, or have developed, within a distributed knowledge infrastructure.

This should also help to identify pathways for full participation of and collaboration with all stakeholder groups in all regions at all stages, from data generation to analysis and application.

The next steps will be to finalise processes, methods and owners for the further development of the database.

This inevitably generates questions which would be useful to follow up under further DiSSCo related work, e.g. in other DiSSCo-linked projects, under the DiSSCo Coordination and Support Office or the DiSSCo General Assembly Task Group. Some such questions are:

- **What stakeholders should be included in this data?** It would not be possible to include all institutions, agencies and projects with an interest in biodiversity or biodiversity data.
- **What information can usefully be captured, and how should this be structured?** Free text information will be difficult to use. As far as possible, the goal should be to collect structured information that supports analysis and reasoning. This probably will require standardized vocabularies to identify classes of data handled by each stakeholder, geographic and taxonomic scope, types of dependency between stakeholders, etc.
- **To what extent should we seek to maintain a current and updated view of this information?** And - maybe even more importantly - how could such a process not only be standardized but also automated so as to make the process efficient and accurate over time?



This work clearly demonstrates the complexity of the landscape; the number of activities, their often-overlapping missions, and their implementation via work programmes on different timelines and at different scales of responsibility. Unless this complexity is understood, there is great risk of inadvertent conflict or duplication of effort. This effort should form a basis for continuation of convergence activities to help identify critical elements that need to be created or sustained and indicate opportunities for better alignment or unification.



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Annex 1

Environmental RIs used in the analysis

In the ESFRI

ACTRIS *Aerosols, Clouds and Trace Gases Research Infrastructure*
AnaEE *Infrastructure for Analysis and Experimentation on Ecosystems*
AQUACOSM *Network of Leading European Aquatic Mesocosm Facilities Connecting Mountains to Oceans from the Arctic to the Mediterranean*
ARISE *Atmospheric Dynamics Research Infrastructure in Europe*
DANUBIUS *International Centre for Advanced Studies on River-Sea Systems*
DiSSCo *Distributed System of Scientific Collections*
EISCAT_3D *Next Generation European Incoherent Scatter Radar System*
eLTER *Integrated European Long-term Ecosystem Research Network*
EMSO *European Multidisciplinary Seafloor and Water-Column Observatory*
EPOS *European Plate Observing System*
EUFAR *European Facility for Airborne Research in Environmental and Geo-Sciences*
Euro-Argo *European Contribution to the International Argo Programme*
EUROFLEETS *New Operational Steps towards an Alliance of European Research Fleets*
EuroGOOS *European Global Ocean Observing System*
GROOM *Gliders for Research, Ocean Observation and Management*
IAGOS *In-service Aircraft for a Global Observing System*
ICOS *Integrated Carbon Observation System*
INTERACT *International Network for Terrestrial Research and Monitoring in the Arctic*
IS-ENES2 *Infrastructure for the European Network for Earth System Modelling*
JERICO *Joint European Research Infrastructure Network for Coastal Observations*
LIFEWATCH *e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research*
SIOS *Svalbard Integrated Arctic Earth Observing System*

External to the ESFRI

AmphibiaWeb
ARIADNE: *Advanced Research Infrastructure for Archaeological Datasets Networking in Europe*
Atlas of Living Australia
BioCAsE: *Biological Collection Access Service*
Biodiversity Heritage Library
Canadian Museum of Nature Online Collection
Catalogue of Life
CERN: *Chinese Ecosystem Research Network*



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



CONABIO: *National Commission for the Knowledge and Use of Biodiversity*
DataONE: *Data Observation Network for Earth*
EMBL-EBI: *The European Molecular Biology Laboratory*
EMBRC-ERIC: *European Marine Biodiversity Resource Centre*
EMSO-ERIC: *The European Multidisciplinary Seafloor and Water-Column Observatory*
Encyclopedia of Life
EU BON: *European Biodiversity Observation Network*
Euro+Med PlantBase
Europeana Collections
GBIF: *Global Biodiversity Information Facility*
GEO: *Group on Earth Observations*
GEOSS: *Global Earth Observation System of Systems*
GGBN: *Global Genome Biodiversity Network*
GRSciColl: *Global Registry of Scientific Collections*
iBOL: *International Barcode of Life*
iDigBio: *Integrated Digitized Biocollections*
iDiv: *German Centre for Integrative Biodiversity Research*
IODP: *International Ocean Discovery Program*
NCBI: *National Center for Biotechnology Information*

Annex 2

GBIF Statement at the inaugural DiSSCo General Assembly

Presented by Executive Secretary Joe Miller

“GBIF is excited to see DiSSCo’s vision and action to increase the rate of biodiversity data mobilization in Europe. European data is key not only to inform us about European biodiversity but due to the large amount of collections from other parts of the world that are held in Europe, digitization here is key to filling global biodiversity knowledge gaps as well. DiSSCo will have a global impact.

The Secretariat urges the DiSSCo National Nodes to work closely with the strong GBIF national nodes in Europe and their regional representatives, which I am happy to see several here today. DiSSCo’s nodes are invited to take part in joint projects involving GBIF nodes and to support joint efforts to engage new data partners and national participation in new countries especially in eastern Europe, the Balkans and Mediterranean where participation has been weak, this is an area where we can work together. The synergy of GBIF and DiSSCo provides an opportunity for nations to access funds to mobilize biodiversity data that are aligned with national, European and global priorities. This is everyone’s primary goal: to provide FAIR data for research and policy decision makers.



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



GBIF applauds the ongoing work in digitization and loans and collection management system developments that will be key to data mobilization success. In the meantime, we encourage DiSSCo partners to continue to mobilize data directly with GBIF during the DiSSCo Prepare phase and we look forward to the transition of data coordination and the data quality and quantity increase that the DiSSCo infrastructure will soon provide.

GBIF fully supports DiSSCo and the proposed convergence to better partner to further develop critical biodiversity knowledge to meet the combined challenges of the biodiversity and climate crises, as well as other complex problems.”



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

