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
MS8.6: Identifying Indicators for Alignment

Author(s)
Lisa French, Eva Alonso, Ana Casino, François Dusoulier, Heli Fitzgerald, Quentin Groom, Aino Juslén, Laurence Livermore, Patricia Mergen, Damian Tonon, Emily Veltjen, Sarah Vincent, Katharine Worley, Vincent Smith

Identifier of the author(s)
Lisa French: https://orcid.org/0000-0001-7279-8582
Eva Alonso: https://orcid.org/0000-0001-5336-9723
Ana Casino: https://orcid.org/0000-0002-9869-6573
François Dusoulier: https://orcid.org/0000-0001-9062-5239
Heli Fitzgerald: https://orcid.org/0000-0002-6754-6409
Quentin Groom: https://orcid.org/0000-0002-0596-5376
Aino Juslén: https://orcid.org/0000-0001-9434-5250
Laurence Livermore: https://orcid.org/0000-0002-7341-1842
Patricia Mergen: https://orcid.org/0000-0003-2848-8231
Emily Veltjen: https://orcid.org/0000-0002-3170-3345
Sarah Vincent: https://orcid.org/0000-0002-4012-0571
Katharine Worley: https://orcid.org/0000-0003-2377-6840
Vincent Smith: https://orcid.org/0000-0001-5297-7452

Affiliation
Lisa French, Laurence Livermore, Vincent Smith, Sarah Vincent: Natural History Museum, London
Eva Alonso: Naturalis Biodiversity Center, Leiden
Ana Casino, Damian Tonon: Consortium of European Taxonomic Facilities
François Dusoulier, Katharine Worley: Muséum national d'Histoire naturelle, Paris
Quentin Groom, Patricia Mergen: Meise Botanic Garden
Aino Juslén: Finnish Museum of Natural History
Emily Veltjen: Instituut voor Natuur- en Bosonderzoek, Brussels

Contributors
Helen Hardy, Ella Howes, Matt Woodburn: Natural History Museum, London
Michel Guiraud, Salomé Landel: Muséum national d'Histoire naturelle, Paris
Dimitris Koureas, Mil de Reus: Naturalis Biodiversity Center, Leiden
Joe Miller: Global Biodiversity Information Facility
Leif Schulman: Syke

Publisher

Identifier of the publisher

Resource ID

Publication year
2021

Related identifiers

Is it the first time you submit this outcome?
Yes

Creation date
23/11/2021

Version

Citation
French, L, Alonso, E., Casino, A., Dusoulier, F., Fitzgerald, H., Groom, Q., Juslén, A., Livermore, L., Mergen, P., Tonon, D., Veltjen, E., Vincent, S., Worley, K. & Smith,V. (2021) Identifying Indicators for Alignment: DiSSCo Prepare WP 8 - Milestone 8.4.

Abstract
The Distributed System of Scientific Collections (DiSSCo) is part of an international landscape of bio-, geodiversity, environmental and life sciences related research infrastructures and organisations. This milestone seeks to provide context on DiSSCo's current positioning within this landscape and to identify opportunities for future alignment and co-operation. The outputs from this report will help to inform future stakeholder engagement plans and prioritisation.

A stakeholder analysis workshop identified the key stakeholders within this domain, with the Global Biodiversity Information Facility (GBIF), Catalogue of Life (CoL) Geoscience Collections Access Service (GeoCASe), Biodiversity Information Standards (TDWG) and the International Barcode of Life all classified as having high influence and interest in DiSSCo.

The stakeholder analysis was supported by insight from the infrastructure contact zones survey, which is an analytic framework used to identify the possible synergies, complementarities and collaboration areas of DiSSCo with other organisations. This provides a systematic and standardised methodology for ranking a wide range of the activities of the different infrastructures relevant to the biodiversity informatics domain. LifeWatch and GBIF both aim to be more generalist infrastructures within the biodiversity informatics landscape, and operate across a large number of activities. There is contact between the activities of these organisations and DiSSCo, and enhanced collaborations may help to maximise the value of projects within this space. DiSSCo will also benefit from the expertise of more specialist infrastructures, such as GeoCASe and the Biodiversity Heritage Library.

Content keywords
organisational

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP8

Project output
Milestone report

Deliverable/milestone number
8.6

Dissemination level
Internal (project partners)

License
CC0 1.0 Universal (CC0 1.0)

Resource type
Text

Format
pdf

Funding Programme
H2020-INFRADEV-2019-2

Contact email
lisa.french@nhm.ac.uk

Rights

DiSSCo PREPARE

H2020-INFRADEV-2019-2
Grant Agreement No 871043

DiSSCo Prepare WP8.3

MS8.6: Identifying Indicators for Alignment

Lisa French, Eva Alonso, Ana Casino, François Dusoulier, Heli Fitzgerald, Quentin Groom, Aino Juslén, Laurence Livermore, Patricia Mergen, Damian Tonon, Emily Veltjen, Sarah Vincent, Katharine Worley, Vincent Smith

NATURAL HISTORY MUSEUM

Abstract

The Distributed System of Scientific Collections (DiSSCo) is part of an international landscape of bio-, geodiversity, environmental and life sciences related research infrastructures and organisations. This milestone seeks to provide context on DiSSCo's current positioning within this landscape and to identify opportunities for future alignment and co-operation. The outputs from this report will help to inform future stakeholder engagement plans and prioritisation.

A stakeholder analysis workshop identified the key stakeholders within this domain, with the Global Biodiversity Information Facility (GBIF), Catalogue of Life (CoL) Geoscience Collections Access Service (GeoCASe), Biodiversity Information Standards (TDWG) and the International Barcode of Life all classified as having high influence and interest in DiSSCo.

The stakeholder analysis was supported by insight from the infrastructure contact zones survey, which is an analytic framework used to identify the possible synergies, complementarities and collaboration areas of DiSSCo with other organisations. This provides a systematic and standardised methodology for ranking a wide range of the activities of the different infrastructures relevant to the biodiversity informatics domain. LifeWatch and GBIF both aim to be more generalist infrastructures within the biodiversity informatics landscape, and operate across a large number of activities. There is contact between the activities of these organisations and DiSSCo, and enhanced collaborations may help to maximise the value of projects within this space. DiSSCo will also benefit from the expertise of more specialist infrastructures, such as GeoCASe and the Biodiversity Heritage Library.

Keywords

Stakeholder, Research Infrastructures, Stakeholder Analysis

Index

01 INTRODUCTION

The Distributed System of Scientific Collections (DiSSCo) will operate within a complex landscape of biodiversity informatics organisations. Successful partnership working with related infrastructures will be an essential element in achieving DiSSCo's mission and strategy. There is a need to identify the areas of contact between infrastructures in this landscape to develop a framework for strategic alignment and co-operation.

This milestone aims to provide an overview of the research infrastructure (RI) landscape in relation to future DiSSCo services and activity, with a focus on European RIs and related international bio/geodiversity organisations (these organisations are referred to as 'infrastructures' throughout the rest of this document). This is achieved through three analyses:

1. **Current Positioning:** The outlines DiSSCo's current position within the European Strategy Forum on Research Infrastructures (ESFRI) landscape, and gives an overview of DiSSCo's current and past collaborative projects.

2. **Infrastructure Stakeholder Analysis:** The influence and interest of each infrastructure stakeholder was mapped on a stakeholder matrix during workshops held in October and November 2021. This analysis will inform future stakeholder engagement plans.

3. **Infrastructure Contact Zones:** The DiSSCo Interim General Assembly commissioned a task force which analysed the complex landscape of biodiversity projects. The dataset from this work has been used in this milestone for a landscape analysis, helping to inform the areas of activity where collaboration between infrastructures is likely to be beneficial.

The outputs from this milestone will be used to develop stakeholder engagement recommendations for the final deliverable for Task 8.3 (T8.3). The stakeholder analysis will also inform work package 4 (WP4), as T4.2 and T4.4 will undertake a set of interviews with research infrastructures in order to identify market opportunities and options to charge for DiSSCo services. The approach taken in this milestone will also be applied to other stakeholder groups, including policy advisory organisations, industrial and commercial partners and DiSSCo's user community.

Task Partners

Natural History Museum, London (NHM)
Consortium of European Taxonomic Facilities (CETAF)
Finnish Museum of Natural History (Luomus)
Meise Botanic Garden (MeiseBG)
Muséum national d'Histoire naturelle (MNHN)

Glossary of Infrastructures

Infrastructure	Acronym	Description/Activity
Atlas of Living Australia	ALA	Australian infrastructure that pulls together Australian biodiversity data.
Analysis and Experimentation on Ecosystems	AnaEE	A distributed infrastructure which will have state of the art facilities to allow experimental manipulations for managed and unmanaged terrestrial and aquatic ecosystems.
BioBanks and BioMolecular resources RI	BBMRI	A research infrastructure for biobanking, aiming to boost biomedical research.
Biodiversity Heritage Library	BHL	Open Access library for biodiversity literature and archives.
Common Language Resources and Technology Infrastructure	CLARIN	A distributed infrastructure that provides access to specialised data collections, knowledge and technical capabilities to conduct research involving the processing of large collections of texts in natural languages.
Catalogue of Life	CoL	Provides a comprehensive index of species.
Digital RI for the Arts and Humanities	DARIAH	A network to support digitally enabled research and teaching in the arts and humanities.
EGI Advanced Computing for Research	EGI	e-Infrastructure which provides advanced computing and data analytics services for research and innovation.
ELIXIR	ELIXIR	Co-ordinates life science resources from across Europe, including databases, software tools, training materials, cloud storage and supercomputers.
Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research	eLTER	Aims to support ecosystem and critical zone research, including socio-ecological research.
European Marine Biological Resource Centre	EMBRC	Aims to advance fundamental and applied marine biology and ecology research.
European Infrastructure for Plant Phenotyping	EMPHASIS	This will provide facilities, resources and services for plant phenotyping.
European Multidisciplinary Seafloor and water column Observatory	EMSO	A consortium with regional facilities observing the oceans, with data and services provided to users.
Environmental Research	ENVRI	A community of Environmental Research

Infrastructure	Acronym	Description/Activity
Infrastructures		Infrastructures, which includes DiSSCo.
European Open Science Cloud	EOSC	Infrastructure providing researchers FAIR data and services - aims to aggregate services provided by several providers.
EOSC-Life	EOSC-Life	This project brings together 13 Life Science ESFRI RIs, and will publish FAIR data and a catalogue of services from the participating RIs for the management, storage and reuse of data in the EOSC.
European RI for Heritage Science	E-RIHS	A distributed infrastructure which gives access to facilities, resources and services in the field of Heritage Science.
EUDAT	EUDAT	An infrastructure which supports research in Europe through integrated data services and resources.
Euro BioImaging	Euro BioImaging	Offers access to imaging technologies, training and data services for biological and biomedical imaging.
European High Performance Computing Joint Undertaking	EURO-HPC	Aims to develop a world class supercomputing ecosystem in Europe.
EU OpenScreen	EU OpenScreen	Provides access to high-throughput screening facilities and medicinal chemistry groups.
Group on Earth Observations	GeoBON	Facilitates the development of a global biodiversity observation network.
Geoscience Collections Access Service	GeoCASe	Data network and web portal for minerals, rocks, meteorites and fossils held in museums and research institutions.
Global Biodiversity Information Facility	GBIF	International network and data infrastructure. Provides data standards and open source tools to enable data-holding institutions to share species occurrence records.
Global Genome Biodiversity Network	GGBN	An international network that aims to help its members make their DNA and tissue collections discoverable for research through a networked community of biodiversity biobanks.
International Barcode of Life	iBOL	A research alliance building DNA barcode reference libraries to aid the discovery and identification of multicellular life.
Integrated Digitized Biocollections	iDigBio	US infrastructure that supports national digitisation efforts.

Infrastructure	Acronym	Description/Activity
iNaturalist	iNaturalist	A social network platform which allows individuals to share biological observations and identifications.
LifeWatch	LifeWatch	A research infrastructure for biodiversity and ecosystem research.
Microbial Resource Research Infrastructure	MIRRI	Brings together microbial domain Biological Resource Centres, culture collections and research institutes to facilitate access to microorganisms and associated data and services.
Partnership for Advanced Computing in Europe	PRACE	Provides access to high performance computing and data management resources and services for scientific and engineering applications.
Biodiversity Information Standards	TDWG	Develops data standards and guidelines for recording data about organisms, including Darwin Core.
World Federation of Culture Collections	WFCC	Concerned with the collection, authentication, maintenance and distribution of cultures of microorganisms and cultured cells.

02 CURRENT POSITIONING

There are a number of related projects which have looked at the research infrastructure landscape. This includes work from a wider ESFRI perspective, and those which have looked at biodiversity informatics organisations more closely related to future DiSSCo services.

ESFRI 2018 Landscape Analysis

ESFRI was formed in 2002 and supports a 'coherent and strategy-led approach to policy-making on research infrastructures in Europe'. It publishes roadmaps for the construction and development of pan-European research infrastructures, with DiSSCo included in the 2018 roadmap. The ESFRI 2018 roadmap included a landscape analysis which outlined the connections between ESFRI RIs, as well as gaps, challenges and future needs.

ESFRI RIs are categorised into six research domains, with DiSSCo classified within the Environment Domain. The Environment Domain is further divided into Atmosphere, Hydrosphere, Biosphere and Geosphere, with DiSSCo within the Biosphere subdomain. DiSSCo will help to address many of the key challenges identified by ESFRI within the Biosphere domain, including addressing the taxonomic gap and monitoring biodiversity and ecosystem change. This will require collaboration with related research infrastructures, including ensuring data is interoperable through the development of data quality, metadata and data preservation standards (ESFRI, 2018). Figure 1 shows the connections between RIs in the Biosphere domain.

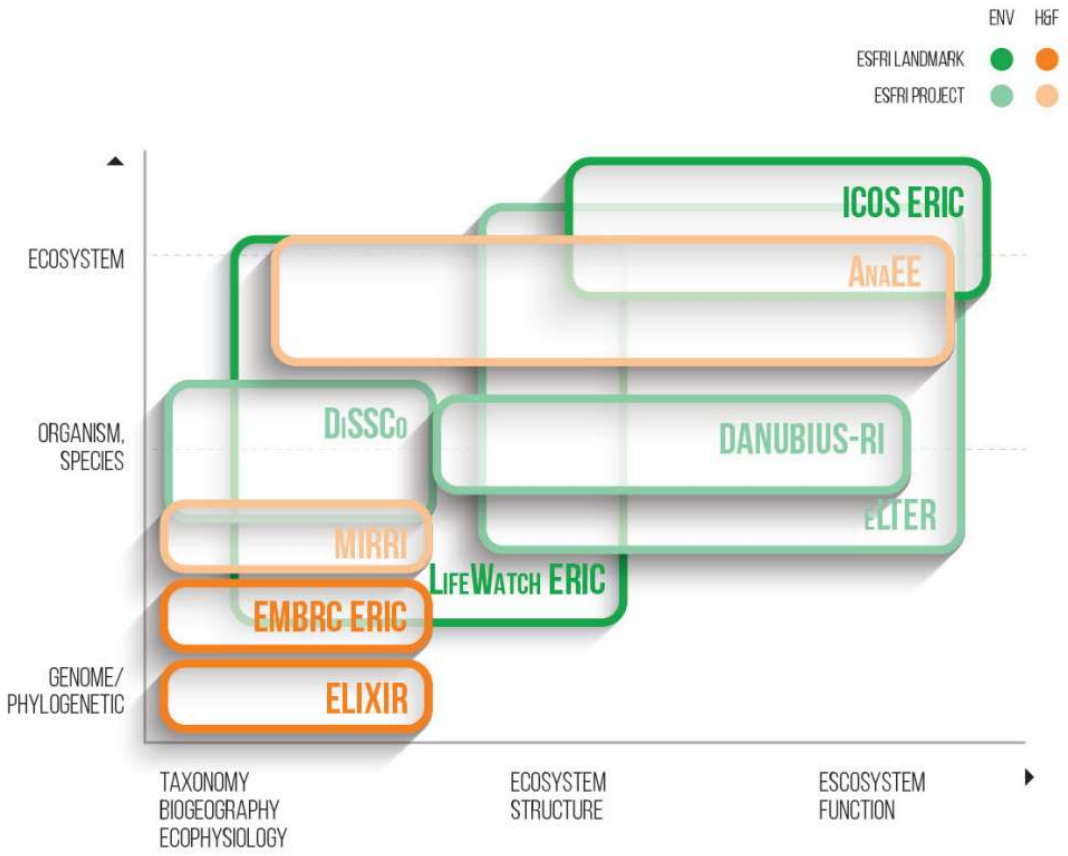

Figure 1: Landscape of biodiversity and ecosystem ESFRI Research Infrastructures Image taken from ESFRI Strategy Report in Research Infrastructures: Roadmap 2018 (ESFRI, 2018).

DiSSCo is positioned by ESFRI as having a research agenda within the “taxonomy, biogeography, ecophysiology” and “organism, species” classifications. MIRRI and LifeWatch are both represented as having some direct overlap with DiSSCo’s research agenda. Other RIs work within similar areas to DiSSCo: DANUBIUS and eLTER are classified in the “organism, species” area, and AnaEE, EMBRC and ELIXIR are within “taxonomy, biogeography, ecophysiology” (ESFRI, 2018). These infrastructures are described in more detail in the glossary above

Within the context of the ESFRI landscape, DiSSCo is a relatively specialised infrastructure, with a tighter research agenda. This is visible in Figure 1 by the smaller area of the rectangle, compared to generalist infrastructures such as LifeWatch and eLTER for example. MIRRI, EMBRC and ELIXIR are also specialised within the ESFRI Biosphere domain, although all three are also classified within the Health & Food domain, with ELIXIR taking a very generalist approach within Health & Food. This analysis can help to inform future collaborative projects, with DiSSCo able to provide expertise within its niche (ESFRI, 2018).

The ESFRI 2018 roadmap also addresses the wider research infrastructure landscape, showing the links between the seven research domains. DiSSCo is identified as having links with the Health & Food Domain, with natural science collections able to help address climate change and food security, and the Digital domain due to DiSSCo’s open data approach (ESFRI, 2018). Although not mentioned in the ESFRI roadmap, DiSSCo could also be

considered to have ties with the Social & Cultural Innovation Domain, in particular with the E-RIHS and DARIAH .

EOSC is recognised by the ESFRI Landscape analysis as being a 'fundamental enabler of the digital transformation of science' which will help to interconnect existing RIs. The report recognises the potential impact of EOSC, and its ability to offer access and reuse of research data (ESFRI, 2018). Since the publication of the 2018 roadmap, EOSC-Life has brought together 13 European life science RIs in order to establish the EOSC as a space for digital biology in Europe. The infrastructures involved include MIRRI, ELIXIR and EMBRC, and it is likely DiSSCo will have interactions with this project in future. EOSC-LIFE aims to publish data resources from life science RIs in the EOSC, and provide policies and guidelines.

The 2018 ESFRI landscape analysis has been used to inform the infrastructure stakeholder analysis described in section 2 of this report. ESFRI will publish the 2021 roadmap in December 2021, which will include an updated landscape analysis.

Prior Work

ICEDIG: The Innovation and consolidation for large scale digitisation of natural heritage (ICEDIG) project aimed to develop a framework for DiSSCo to identify international stakeholders and related research activities.This project undertook a landscape analysis to map domain specific infrastructures within bio and geo-diversity, as well as related environmental research infrastructures and e-Infrastructures (such as the EOSC) (Smith & Goodson, 2019).

A centralised database was created which documented projects, initiatives, platforms and portals with a sample dataset for 35 infrastructures. The project developed a prototype dashboard which visualised the collaborative projects between RIs, the services offered by RIs and the geographical location of these infrastructures (Figure 2) . This work demonstrated the complexity of the landscape, and is a starting point for considering activities where alignment and collaboration would be beneficial (Goodson *et al,* 2020).

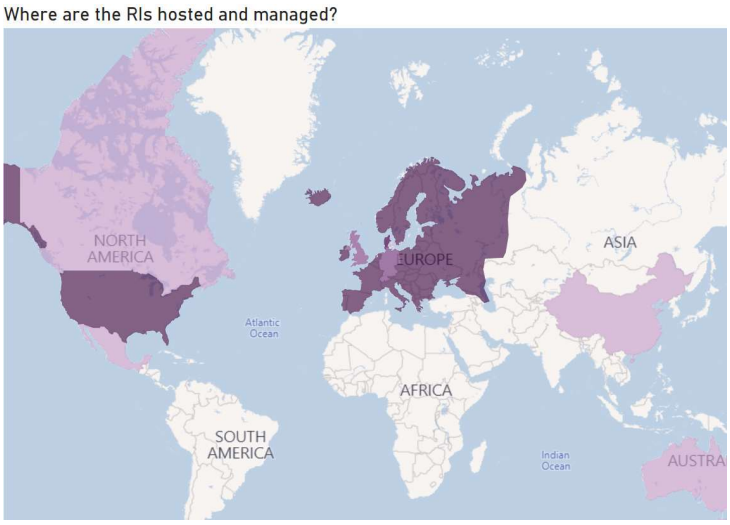

Figure 2: Visualisation from the prototype research infrastructure dashboard. This shows the geographical locations from the sample dataset, with a darker colour indicating that there are a higher number of infrastructures in the area (Goodson et al, 2020).

Current Collaborative Projects

DiSSCo has entered into a number of collaborative projects, which help to support the goals and development of DiSSCo. These projects also help to strengthen relationships with relevant stakeholder infrastructures, including developing standards, policies and technical infrastructure. The infrastructures involved in these projects can be seen in Figure 3.

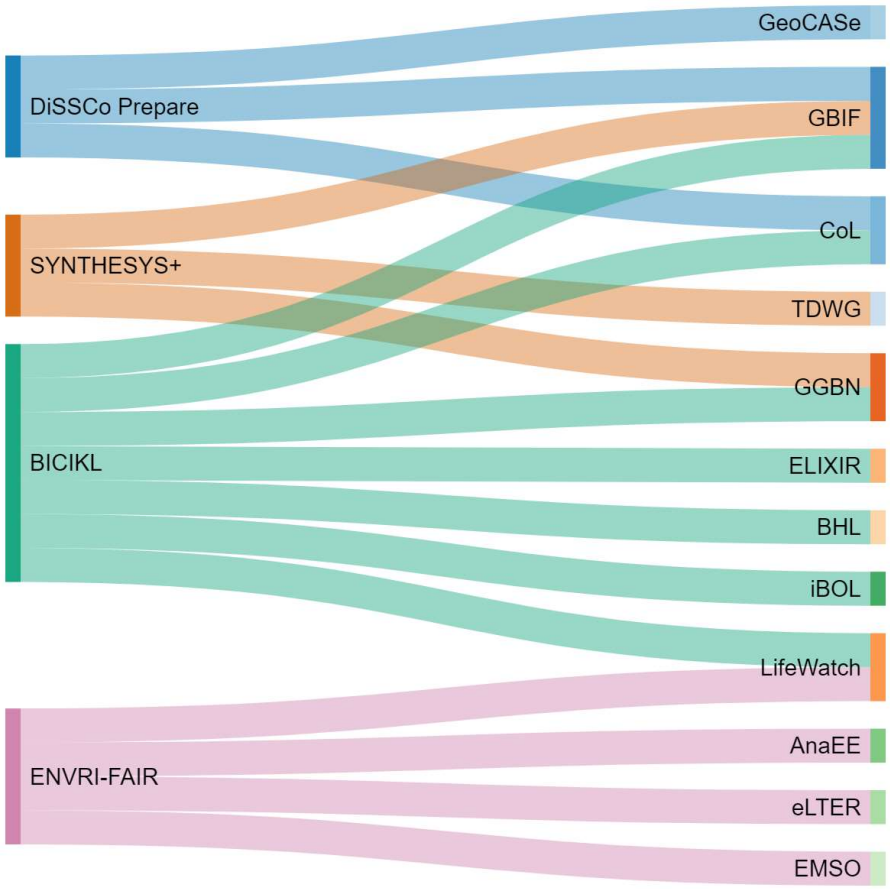

Diagram created using SankeyMATIC.

Figure 3: A visualisation showing the infrastructures which are involved in collaborative projects with DiSSCo. The left-hand side shows the European Commission (EC) funded projects, and the right hand side lists the infrastructures. ENVRI-FAIR only shows the infrastructures which have been identified as a DiSSCo stakeholder (Figure 6), rather than all project members. Each project is described in more detail below, and the glossary gives a description of each infrastructure.

DiSSCo Prepare: H2020-INFRADEV-02-2019-2020. Project Number: 871043 - The digital unification of Europe's natural science assets.

DiSSCo Prepare (DPP) is the preparatory project phase of the DiSSCo-RI, which will develop a Construction Masterplan across five areas: technical, scientific, data, organisational and financial readiness. DPP is organised into nine work packages. GeoCASe and CoL are key partners in T5.4, which will develop a plan for the improvement of the technical infrastructure in geo-collection data and taxonomic services. GBIF is involved in T6.4, which links the design of DiSSCo technical architecture and service provision to the European and global technology landscape.

SYNTHESYS+: H2020-INFRAIA-2018-1. Project Number: 823827 - A digital access boost for natural history collections.
This project aims to create a high quality approach to the management, preservation and access to European natural history collections, and will lay the foundations for DiSSCo by creating an accessible, integrated European resource for research users in the natural sciences. It is the fourth and final iteration of SYNTHESYS programme, which has developed a collections infrastructure with international partners. SYNTHESYS+ includes access, research and networking activities. The Global Genome Biodiversity Network (GGBN) leads a work package to develop, implement and disseminate standardised best practices to support sequencing and biobanking. Biodiversity Information Standards (TDWG) leads a task related to digital standards and processes, and GBIF is developing an international roadmap for biodiversity infrastructure component areas.

BICIKL (Biodiversity Community Integrated Knowledge Library): H2020-INFRAIA-2020-1. Project Number: 101007492 - Advanced data library for the biodiversity field.
This project aims to make data interoperable across infrastructures for biodiversity research and knowledge sharing. This includes links to molecular sequence (European Nucleotide Archive, ELIXIR), specimens (MeiseBG, Naturalis, Botanic Garden and Botanic Museum, Berlin), biodiversity literature (Plazi) and taxonomic names (CoL). In making links and improving the modes of access to these data to make them accessible BiCIKL aims to enable new forms of data analysis and automated workflows that bring together infrastructures in novel ways. DiSSCo is represented by its coordinating institution and several institutional partners as it is not yet a legal entity.

ENVRI-FAIR: H2020-INFRAEOSC-04-2018. Project Number: 824068 - ENVironmental Research Infrastructures building Fair services Accessible for society, innovation and research.
ENVRI-FAIR is a project part of the EOSC thematic clusters. It aims to connect environmental RIs. DiSSCo is involved through its coordinator and CETAF, as the proposal was about to be submitted when DiSSCo was officially approved on the ESFRI roadmap. ENVRI-FAIR allows DiSSCo to remain informed and to contribute towards collaborations plans and implementations of ESFRI environmental infrastructures. EOSC-Life was funded under the same program for synergies among ESFRI Life Science Infrastructures and both cluster projects work closely together.

03 INFRASTRUCTURE STAKEHOLDER ANALYSIS

The task partners in DPP8.3 undertook a stakeholder analysis exercise of key research infrastructures and biodiversity informatics organisations. This exercise was intended to help identify the key actors within this landscape, and to assess their relative interest and influence to the DiSSCo RI. A stakeholder influence-interest matrix, which quantifies the influence and interest of a stakeholder from low to high (Figure 4) was used as the primary tool to map and classify stakeholders. This work is intended to help inform future strategic planning, detect potential project risks, and stakeholder engagement approaches of DiSSCo.

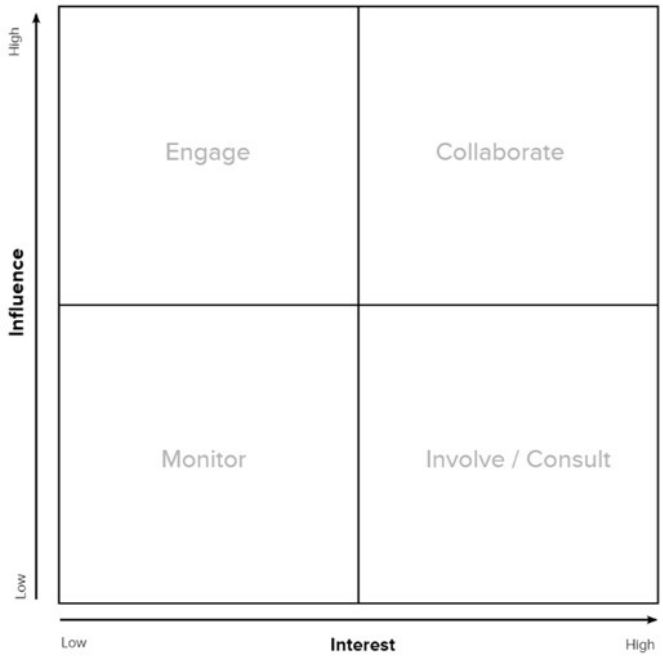

Figure 4: Influence-Interest stakeholder matrix

5.1 Virtual Workshops

Workshop 1: A virtual workshop was held on 25 October 2021, with 13 participants from NHM, Luomus, CETAF, Naturalis Biodiversity Center, MNHN and MeiseBG. Participants had a variety of experience working on DiSSCo-related projects, and working with infrastructure stakeholders.

Participants were split into two break out groups, and were asked to map infrastructures against a stakeholder matrix, based on their relevance to DiSSCo (interest) and their degree of activity and importance in their sector (influence). MURAL, a digital collaboration tool, was used to assign infrastructure stakeholders to an influence-interest matrix (Figure 4).

A list of infrastructure stakeholders was prepared in advance of the workshop by task members from MNHN, CETAF and NHM. Each break out group was asked to assign a core set of infrastructures, which included the infrastructures involved in the contact zones analysis (see section 3), as well as those that currently collaborate with DiSSCo. A list of related infrastructures from the ESFRI roadmap were also provided on the MURAL, allowing participants to assign these infrastructures onto the matrix (Figure 5). Participants were also invited to suggest other infrastructures to include on the stakeholder map, which resulted in the inclusion of GeoBON, ALA and EOSC-Life.

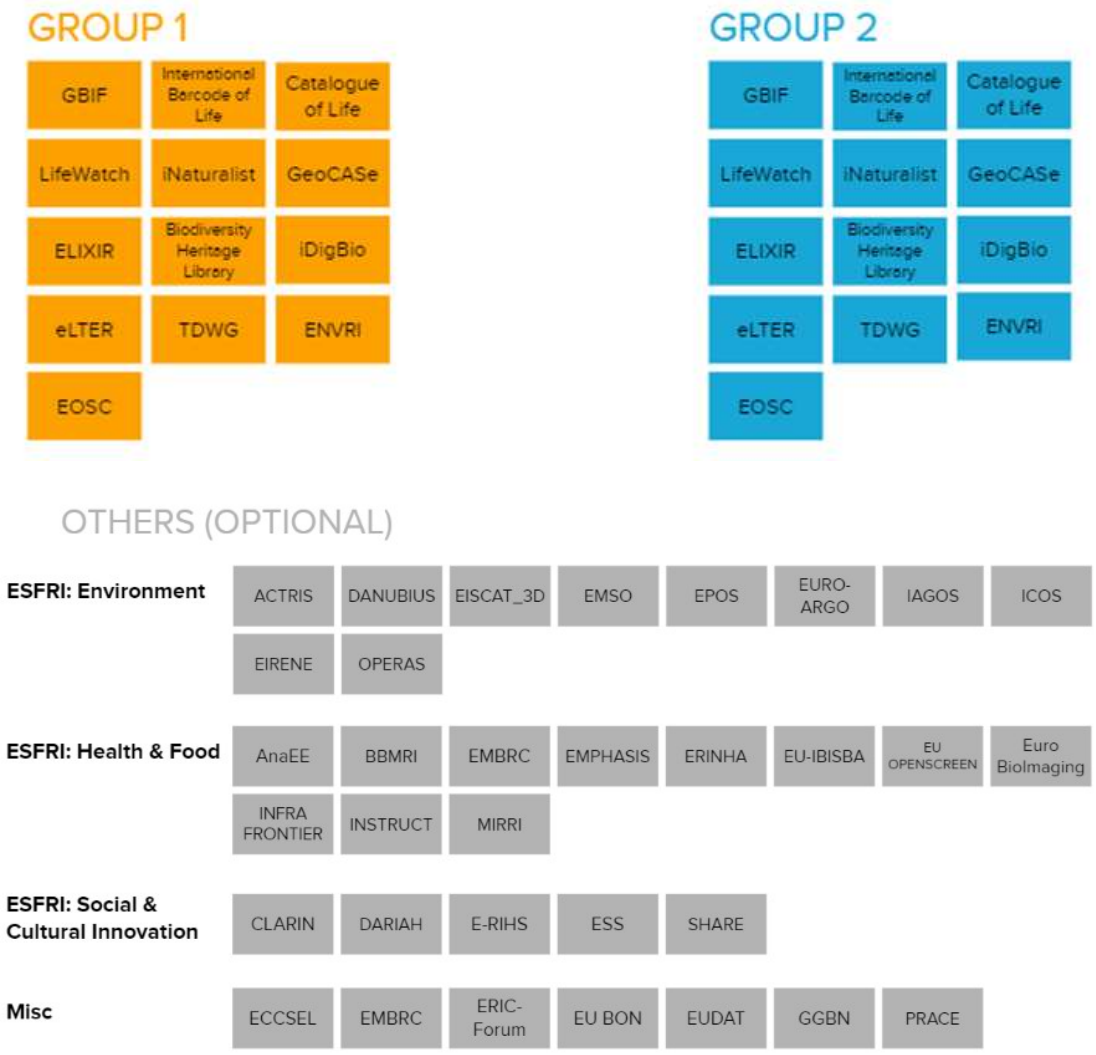

Figure 5: Set of infrastructures considered by the breakout groups. Both groups were asked to place all infrastructures in orange/blue, and could optionally add infrastructures listed under 'Others'. Participants were able to suggest other infrastructures for inclusion on the stakeholder matrix.

All participants then returned to the main plenary to discuss the differences in stakeholder placement between the two groups in order to agree on their final position. In some cases this led to infrastructures being placed between two quadrants.

Workshop 2: A second, shorter, workshop was held on 4th November 2021.This was attended by eight participants from MNHN, NHM, MeiseBG, Luomus, Naturalis and CETAF. Participants were asked to classify stakeholders by whether they were primary (directly benefit from DiSSCo), secondary (indirectly benefit from DiSSCo) or tertiary (no benefit but have influence). They were also asked to consider whether any of these stakeholders were likely to change in influence or interest in future. Participants were also given an opportunity to reflect on the final stakeholder map, and made some minor amendments to the positioning of each infrastructure based on discussions throughout the workshops.

5.2 Stakeholder Map

The research infrastructure and biodiversity informatics organisation stakeholder map can be seen in Figure 6. The infrastructures within each quadrant are discussed in more detail in this section, along with highlights from the conversations in the workshops. Each stakeholder is classified by their influence and interest level in DiSSCo, and whether they are primary, secondary or tertiary stakeholders.

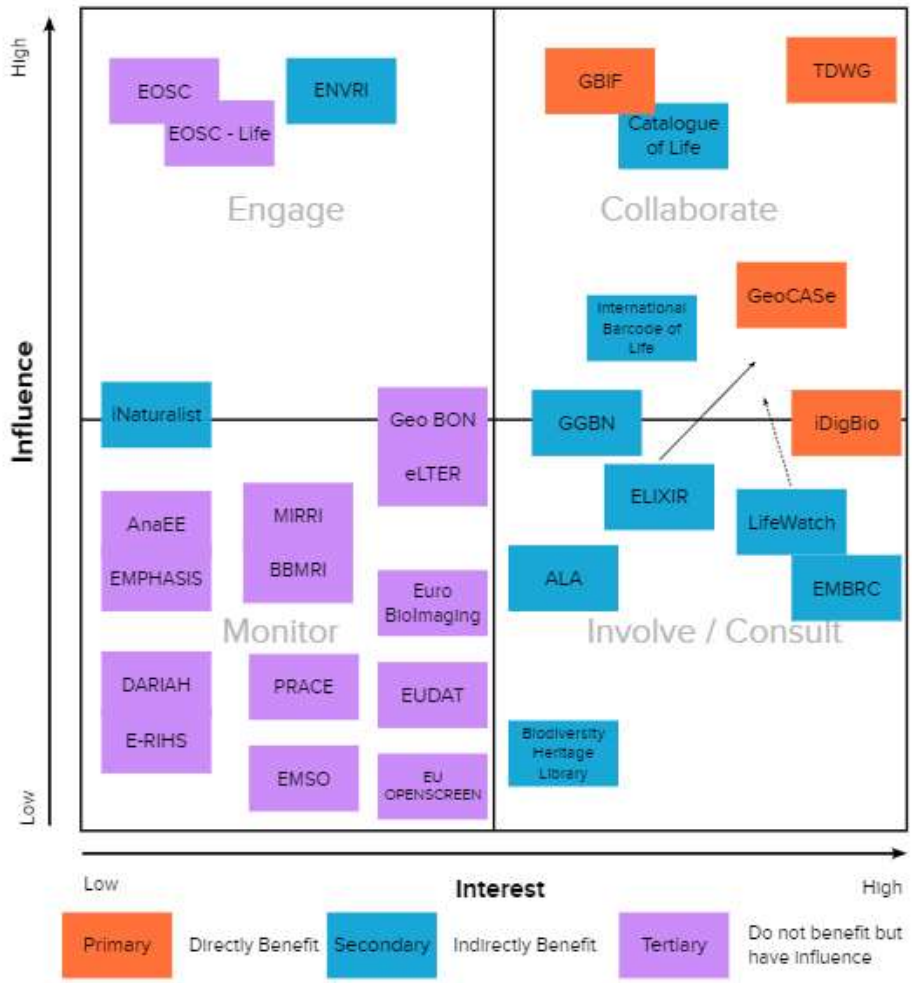

Figure 6: DiSSCo Research Infrastructure Stakeholder Map. Stakeholders were classified based on their level of influence and interest (indicated by the position on the map), and whether they were primary, secondary or tertiary stakeholders (indicated by colour). The arrows for ELIXIR and LifeWatch indicate these infrastructures are considered likely to have more influence in future.

Collaborate

Table 2 provides a list of the infrastructures which fall into the High Influence, High Interest quadrant. These infrastructures are those where collaborative partnership working is likely to be highly beneficial, and DiSSCo already collaborates with most of the infrastructures that fall into this category.

Workshop discussion: Collaboration and engagement between GBIF and DiSSCo was considered important, with GBIF acting as an aggregator for international biodiversity data. DiSSCo is likely to directly benefit GBIF, GeoCASe and TDWG, through interoperability and standards development, and these were classified as primary stakeholders. There was less certainty about the classification of the Catalogue of Life, which has significant links to GBIF, although ultimately it was determined to indirectly benefit (secondary stakeholder) from DiSSCo through its relationship with GBIF.

Table 2: Infrastructures within the high interest, high influence category. Italics indicates the infrastructure was classified as between high influence/low influence.

Stakeholder	**Classification**
GBIF	Primary
TDWG	Primary
GeoCASe	Primary
iDigBio	*Primary*
CoL	Secondary
iBOL	Secondary
GGBN	*Secondary*

Engage

Table 3 describes the infrastructures that fall into the high influence, low interest quadrant, and are infrastructures that DiSSCo is advised to engage with during its construction and development.

Workshop Discussion: EOSC-Life was an additional infrastructure added to the stakeholder map, and may be one of the infrastructures along with ENVRI-FAIR through which DiSSCo can engage with EOSC.

Table 3: Infrastructures within the low interest, high influence category. Italics indicates the infrastructure was classified as between high influence/low influence

Stakeholder	**Classification**
ENVRI	Secondary
iNaturalist	*Secondary*
EOSC	Tertiary
EOSC-Life	Tertiary
GeoBON	*Tertiary*

Consult

Table 4 shows the infrastructures classified in the low influence, high interest quadrant. It may be beneficial for DiSSCo to consult with these infrastructures during its development.

Workshop Discussion: ELIXIR and LifeWatch were both classified as infrastructures which may have more influence on DiSSCo activities in future, with an arrow indicating they were likely to move into the high influence box. Collaborative working with these infrastructures was likely to increase through the BICIKL project.

Table 4: Infrastructures within the high interest, low influence category. Italics indicates the infrastructure was classified as between high influence/low influence.

Stakeholder	**Classification**
iDigBio	*Primary*
ELIXIR	Secondary
ALA	Secondary
BHL	Secondary
LifeWatch	Secondary
EMBRC	Secondary
GGBN	*Secondary*

Monitor

Table 5 shows the infrastructures within the low influence, low interest group. DiSSCo may want to monitor the activities and outputs of these infrastructures.

Workshop Discussion: A number of the infrastructures within this quadrant were considered to be linked, either through collaboration or through scope/research area, which may influence the communication strategy for these infrastructures. This is visualised by clustering the infrastructures on the stakeholder map (Figure 6):

- GeoBON and eLTER: Both GeoBON and eLTER collect metadata about observations in areas that respect specific protocols, notably have time series of measurements and/or observations. eLTER is the European contribution to the international network iLTER which has long term collaborations with GeoBON.
- DARIAH and E-RIHS are closely related to art, humanities, Heritage Sciences and socio-cultural data. CLARIN was not included on the stakeholder map, but it was noted that they are linked with DARIAH and E-RIHS. In several countries, DARIAH and CLARIN have merged ('CLARIAH') but they do operate separately in others.
- AnaEE and EMPHASIS are analysing ecological and biological parameters in controlled *in situ* and *ex situ* environments. AnaEE has a broader ecosystem scope while EMPHASIS focuses specifically on plant phenotyping.
- BMRI and MIRRI: MIRRI brings together microbial domain Biological Resource Centres, culture collections and research institutes, with a focus on Health and Food, Agro-Food and Environment and Energy. BBMRI is an RI for biobanking, with a focus on human health and human cells. There is some linkage between the two RIs, as BBMRI has an interest in human pathogens.

- EUDAT and EOSC: It is likely DiSSCo would directly engage with EOSC. Long-term standing networks such as EUDAT, EGI, EURO-HPC led to the creation of EOSC by the European Commission, and collaborations are already in place. Future DiSSCo services will need to consider EOSC onboarding rules so they can be visible on the EOSC services portal, which has a global, multidisciplinary scope. In terms of reaching out to industrial partners the newly formed GAIA-X network should be considered, which is a Federated and Secure Data Infrastructure, aiming principally at the collaboration with the private sector.

Table 5: Infrastructures within the low interest, low influence category.
Italics indicates the infrastructure was classified as between high influence/low influence.

Stakeholder	**Classification**
EMSO	Tertiary
E-RIHS	Tertiary
eLTER	Tertiary
AnaEE	Tertiary
EMPHASIS	Tertiary
MIRRI	Tertiary
BBMRI	Tertiary
DARIAH	Tertiary
PRACE	Tertiary
Euro BioImaging	Tertiary
EUDAT	Tertiary
EU OpenScreen	Tertiary
iNaturalist	*Secondary*
GeoBON	*Tertiary*

Next Steps

A stakeholder engagement plan will be developed for the deliverable for T8.3, and this stakeholder mapping exercise will help to inform this plan. Participants in the workshop agreed that prioritisation of stakeholder engagement would be best placed once stakeholder mapping had taken place for all stakeholder groups, including industry, DiSSCo's user community and policy advisory bodies.

04 INFRASTRUCTURE CONTACT ZONES

Overview

In February 2020, the DiSSCo Interim General Assembly commissioned a task force to examine the alignment of DiSSCo to related biodiversity informatics organisations and research infrastructures. The goal was to articulate the strategic position of DiSSCo alongside related infrastructures.

The task force developed an analytical framework to analyse the 'contact zones' of DiSSCo against these infrastructures, known as the Infrastructure Contact Zones analysis. The contributions of each member of the task force can be found under the 'Additional Contributors' section of this milestone. A quantitative survey was created to characterise the current and planned activities of major biodiversity informatics organisations (Appendix 1). This framework was inspired by work published by the Finnish Biodiversity Information Facility (FinBIF), which modelled several biodiversity data infrastructures based on their data type and data life cycle phases (Schulman *et al.,* 2021).

Contact between infrastructure organisations was captured across five categories (data, standards, software, hardware and policy), nine types of data (specimens, collection descriptions, opportunistic observations, systematic observations, taxonomies, traits, geological data, molecular data, and literature), and seven phases of activity (creation, aggregation, access, annotation, interlinkage, analysis, and synthesis). These are collectively referred to as an activity in this document, and definitions for each activity can be found in Appendix 1.

GBIF, iBOL, CoL, iNaturalist, BHL, GeoCASe, LifeWatch, eLTER and ELIXIR all agreed to take part in the survey, with each infrastructure scoring their current and planned ambition in each area. They were asked to score their activities on a maturity index, which ranged from no activity (0) to a predominant level (4) (Table 6). This generated a dataset of 6,300 observations, which will be used by the task force to explore organisational niches, overlaps and gaps.

Areas of 'contact' with DiSSCo activities is particularly relevant to the stakeholder analysis in Task 8.3. The contact zones dataset allows for a landscape analysis, identifying the current and future strategic positioning of infrastructures in order to highlight areas for potential future collaboration

Table 6: Shows the definition for each maturity index level in the infrastructure contact zones analysis. Each infrastructure was asked to score their activities against these maturity levels.

Maturity Index Level	**Definition**
P0: No activity/inapplicable	No current/planned activity or inapplicable to an organisation's operations.
P1: Planned	Named a strategy, roadmap or outline development as a proof of concept (evidenced through documentation or a prototype solution).
P2: Presence	Addresses part of the domain/problem set served, sometimes as a dependency to addressing other issues, and in use (evidenced through the use of the solution beyond the developing organisation).

P3: Performance	Addresses a majority/full scope of the domain it serves and in widespread use (evidenced through the richness of feature set and widespread use).
P4: Predominance	A domain leader to which all other innovators would aspire to or work with, addressing the full scope of the domain and sustained through continuous improvement (evidenced through market share).

Landscape Analysis

The contact zones dataset allows for an exploration of the organisational niches that each infrastructure occupies. In the context of the infrastructure contact zones analysis, an organisational niche can be considered as the areas of activity the organisation plans to specialise in. The niches each participating biodiversity informatics organisation plans to occupy at a predominant or performant (maturity index 3 or 4) is visualised in Figure 7.

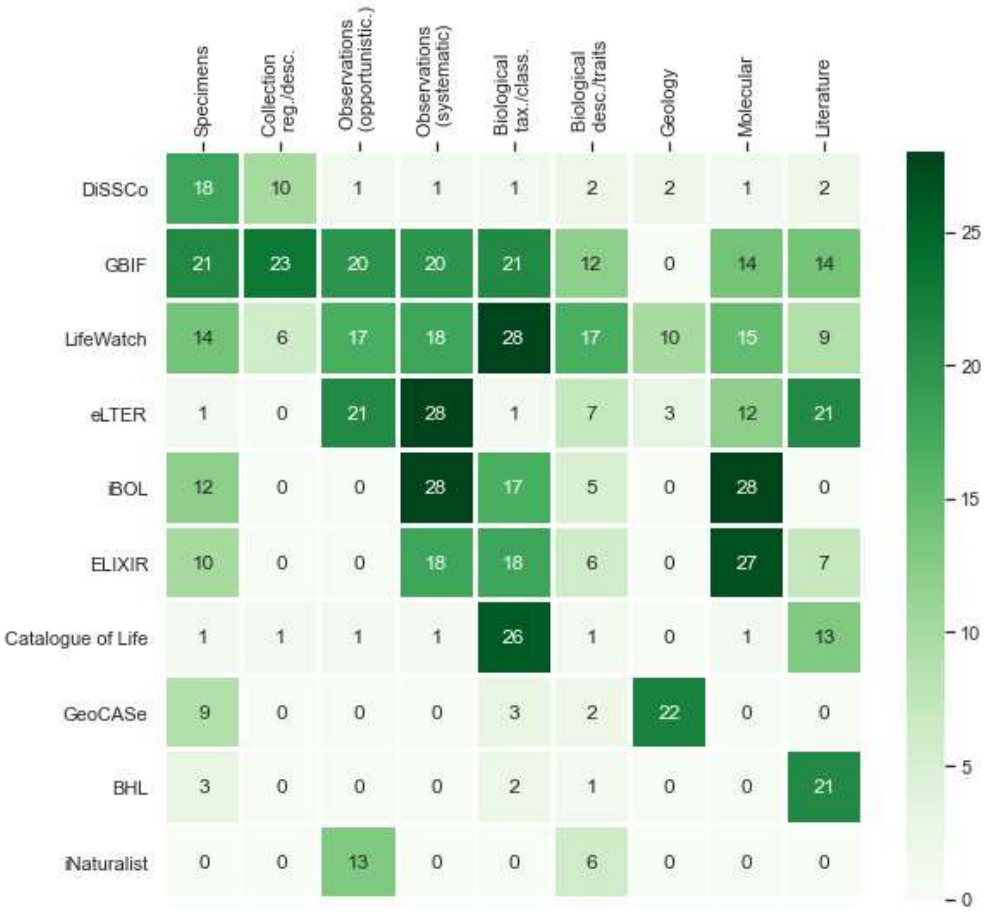

Figure 7: **Infrastructure Niches:** This heatmap shows the number of categories each infrastructure declared their planned ambition to be at maturity index level 3 or 4, grouped by

data type ('scope' of activity), with darker colours representing increasing levels of intense activity. The numbers in each box represents the activities scored at a maturity index 3 or 4, with 35 the maximum number. This figure does not include the Hardware category. Definitions can be found in Table 7 and in Appendix 1.

Some infrastructures are generalists, aiming to have their activities cross a large range of areas. This includes GBIF, with activities scored in all data types apart from Geology (they however deal with Palaeontology, but not with Minerals, Rocks or Meteorites, so Geology has to be understood here excluding Palaeontology). LifeWatch is also a generalist, with activities scored at a high maturity level in all data types, but with a higher area of focus in the biological taxonomy/classification data type.

Others are specialists: with BHL focusing activity mainly in literature, Catalogue of Life in biological taxonomy/classification, GeoCASe in geology and iNaturalist in opportunistic observations. DiSSCo is also somewhat of a specialist, with most of its activity within the specimen and collections registry/description area.

The infrastructure contact zones dataset allows for exploration of areas of contact and synergy between infrastructures, with an area of contact defined as an activity which two or more infrastructures share. Contact between infrastructures does not necessarily indicate that there is duplication of activity, as each infrastructure is likely to take a unique approach to solving a problem. Contact between infrastructures may suggest that there is an opportunity for alignment and co-operation, as collaborative working may help to maximise the value of these projects.

Figure 8 shows which infrastructures share similar ambitions in activities that DiSSCo rated at a performant or predominant level. GBIF and LifeWatch both show a high level of contact with DiSSCo activity. This may be a result of the more 'generalist' nature of these infrastructures, and this is explored further in the next section.

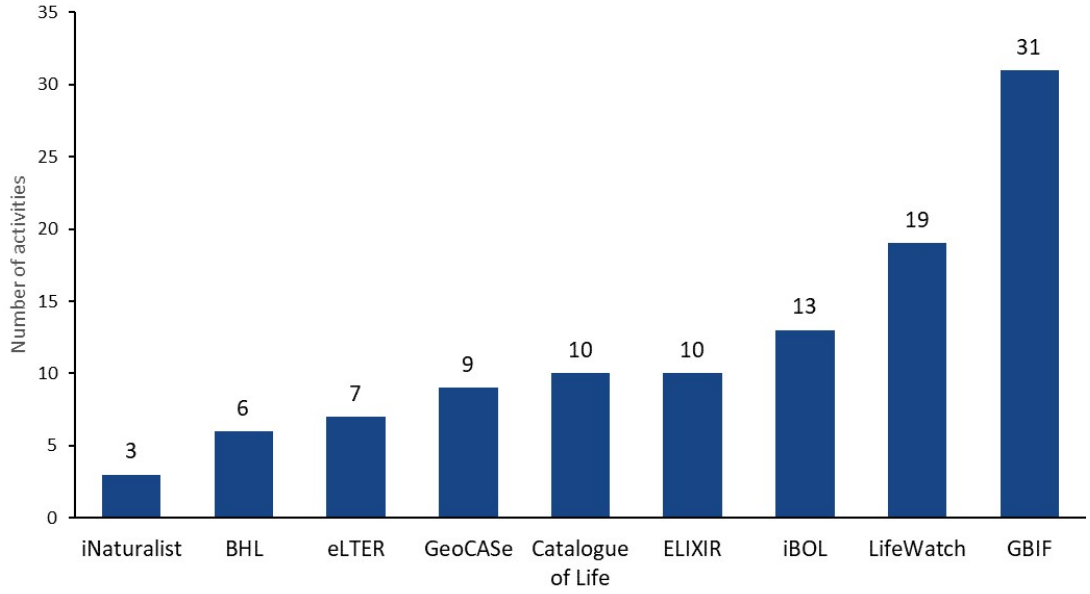

Figure 8: This shows the number of areas where each infrastructure has rated their future 'ambition' maturity index level at 3 or 4, in activities where DiSSCo also aims to be at levels 3 or 4.

Generalists

GBIF and LifeWatch are the two infrastructures with the highest areas of contact with DiSSCo activity as illustrated by Figure 7 and Figure 8. Similarly, Figure 9 shows that LifeWatch and GBIF, labelled to be generalists, have ambitions to execute many activities, visible by the large radius and roughly 150 areas of activity rated at maturity level 3 or 4, in comparison to DiSSCo, which we label to be more of a specialist in comparison, which has 'only' 39 activities rated at this level.

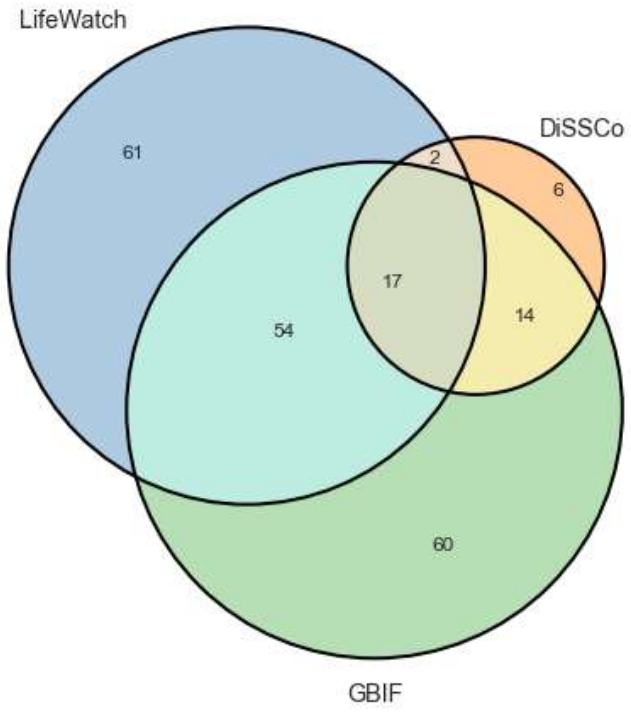

Figure 9: Venn Diagram showing areas of activity shared by LifeWatch, GBIF and DiSSCo The bubble size for each infrastructure represents the number of activities with the ambition to be at maturity level 3 or 4, with the numbers showing the amount of activity in each space. This figure does not include the Hardware category. Table 7 shows the definitions for each maturity level.

The specific areas of GBIF and LifeWatch activity which have contact with DiSSCo activity can be seen in Figure 10. DiSSCo aims to specialise in specimens and collections, visible in Figure 10 by the largest proportion and number of activities in the 'Category' column. Subsequently most overlap or contact with LifeWatch and GBIF falls within this Category. This suggests there are opportunities for GBIF, DiSSCo and LifeWatch to work together to complement each other in reaching common goals.

DiSSCo has seventeen areas of activity it shares with both GBIF and LifeWatch, and fourteen areas it shares only with GBIF. There are only two areas of contact between DiSSCo and LifeWatch which exclude GBIF. This suggests that collaborative activity and partnership working with LifeWatch would often be most beneficial where it is a joint enterprise with GBIF.

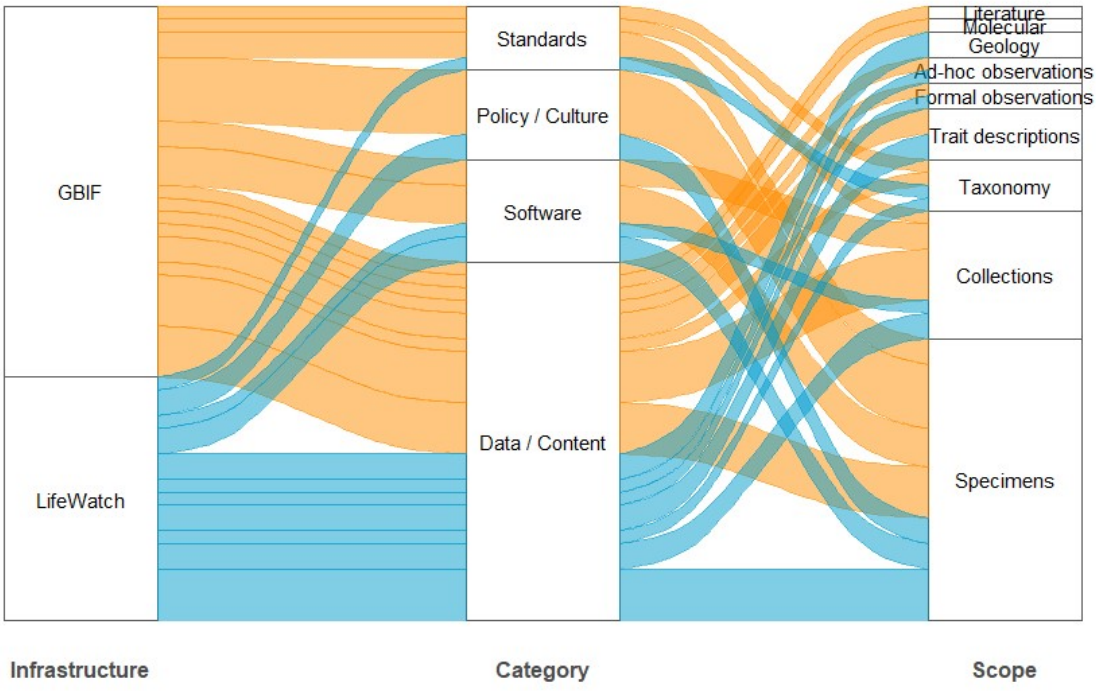

Figure 10: A Sankey diagram showing the activities which LifeWatch and GBIF share with DiSSCo, and does not include activities with no contact with DiSSCo. The thickness of each line represents the number of activities rated at maturity index level 3 or 4 that overlap with DiSSCo. LifeWatch activities are in blue, GBIF in orange. The most common areas of overlap are within the Data/Content category, and the Specimen and Collections scope.

DiSSCo and LifeWatch are relatively new infrastructures and are in the earlier stages of their development, whereas GBIF is much more established. The contact shown in these diagrams represent the ambitions of these infrastructures, rather than the current status, and this may shift in future. This analysis can inform DiSSCo's strategic planning going forward, and should continue to be monitored during the construction and operational phases of DiSSCo.

Specialists

Figure 7 shows that iNaturalist, BHL, Catalogue of Life and GeoCASe all aim to specialise within a specific scope. These infrastructures will be valuable for DiSSCo to consult when working on projects that interlink with these areas of specialism. For example, GeoCASe has 12 activities where it aims to be the only infrastructure at a predominant or performant level, with all of these activities falling within Geology. DiSSCo aims to have a presence in some of these areas, and would therefore benefit from collaboration with GeoCASe.

Next Steps

A dashboard is currently under development by the contact zones task force which will allow exploration of the contact zones dataset (Figure 11), and much of the analysis from this report is informed by the dashboard development.The dashboard will the user to interrogate the dataset in more detail, including organisational niches, possible gaps in activity and the level of contact between all infrastructures.

The contact zones taskforce plan to use this dataset in two publications, one of which will publish the data and another that will provide a high level synthesis. The dataset will also inform the construction plans for improvement of the technical infrastructure in geo-collection data and taxonomic services as part of T5.4.

The contact zones methodology could also be extended to include additional infrastructures, for example TDWG to better cover the standards aspects. The stakeholder analysis may also help to guide which infrastructures could be invited to complete the survey.

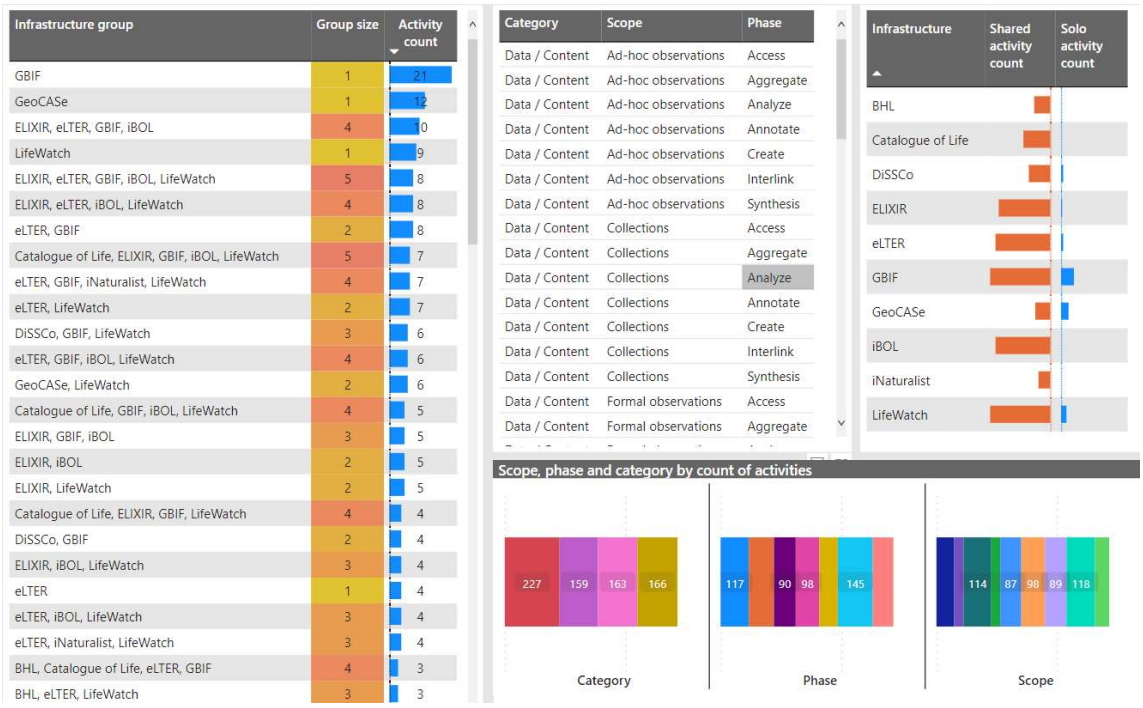

Figure 11: A draft view from the Contact Zones Dashboard

05 CONCLUSIONS

This milestone identified and classified the research infrastructure stakeholders for DiSSCo, and explored the breadth of activities where DiSSCo has contacts with nine other biodiversity informatics organisations.

This task has identified 30 infrastructure stakeholders, with each infrastructure classified on the basis of their interest and influence on DiSSCo activities. The stakeholder matrix suggests an approach to engagement with these organisations: whether to collaborate, to consult, to engage or to monitor (Figure 6). This is a preliminary classification only, and will be used to inform a more detailed stakeholder engagement plan, which will be expanded upon in the deliverable for this task. The infrastructure stakeholder map will continue to be reviewed, as there may be developments which impact upon the positioning of each infrastructure, as well as the potential inclusion of additional infrastructures.

The contact zones dataset has provided a unique opportunity to analyse the biodiversity informatics landscape. This has helped to identify the areas of contact DiSSCo has with other organisations. This can help inform the activities where collaboration is likely to be particularly beneficial. It is important DiSSCo works closely with organisations where there is

a high level of contact. The collaborative partnerships with organisations with a high level of contact will bring together their expertise and unique approaches to address key challenges.

This milestone focussed on DiSSCo's infrastructure stakeholders. Other stakeholder groups, including relevant policy advisory organisations, industrial and commercial partners, and DiSSCo's user community will be investigated in more detail. Together, these analyses will inform a stakeholder engagement plan which will be included in the deliverable for this task.

06 AUTHOR CONTRIBUTIONS

Contribution types are drawn from CRediT - Contributor Roles Taxonomy

Conceptualisation, Writing - original draft: Lisa French

Formal Analysis, Visualization: Lisa French, Vince Smith, Sarah Vincent

Investigation: Eva Alonso, Ana Casino, François Dusoulier, Heli Fitzgerald, Lisa French, Quentin Groom, Aino Juslén, Laurence Livermore, Patricia Mergen, Damian Tonon, Vince Smith, Katharine Worley

Methodology: Lisa French, Vince Smith, Damian Tonon, Katharine Worley

Writing - review & editing: Quentin Groom, Patricia Mergen, Vince Smith, Emily Veltjen, Katharine Worley

Additional Contributors

Investigation: Michel Guiraud, Salomé Landel

Writing - review & editing: Helen Hardy, Ella Howes

Contact Zones Task Force:

Conceptualisation: Ana Casino, François Dusoulier, Aino Juslén, Dimitris Koureas, Patricia Mergen, Joe Miller, Leif Schulman, Vince Smith

Methodology: Leif Schulman, Vince Smith, Matt Woodburn, Sarah Vincent

Investigation: Aino Juslén, Patricia Mergen, Vince Smith, Matt Woodburn

Project Administration: Mil de Reus

Data Curation, Formal Analysis, Visualization: Lisa French, Vince Smith, Matt Woodburn, Sarah Vincent

07 GLOSSARY

Please see the introduction for a description of each infrastructure mentioned in this report.

BiCIKL (Biodiversity Community Integrated Knowledge Library): BiCIKL will build a new European starting community of RIs in biodiversity and life sciences. It will establish open science practices through provision of access to data, associated tools and services.

DiSSCo Prepare: The preparatory phase project for DiSSCo. Preparatory phases of ESFRI roadmap projects aim to bring the new research infrastructure to the required legal, financial and technical maturity level for implementation. DPP is organised into nine work packages, each of which contribute to the implementation of DiSSCo.

ENVRIplus: ENVRIplus was a Horizon 2020 project that aimed to create an interdisciplinary and interoperable cluster of Environmental RIs. The final report was published in October 2019.

GAIA-X: GAIA-X is a project including business, science and political representatives which aims to create a federated and secure data infrastructure.

ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage): This was an EU funded project which supported the implementation phase of DiSSCo, and designed some of the technical, financial, policy and governance aspect required to operate DiSSCo.

Infrastructure: Infrastructures, organisations networks and associations within the biodiversity and environmental research landscape. This does not include individual natural science collection holding institutions.

Primary Stakeholder: Stakeholders who directly benefit from a project.

Secondary Stakeholder: Stakeholders who indirectly benefit from a project

Stakeholder Engagement Plan: A strategy that outlines how project stakeholders will be communicated with in order to achieve project goals.

SYNTHESYS+: SYNTHESYS+ is a Horizon 2020 project which will create an accessible, integrated European resource for research users in the natural sciences. It has access, research and networking streams.

Tertiary Stakeholder: Stakeholders who have no direct or indirect benefit from a project, but do have influence on the project.

08 REFERENCES

ESFRI (2018) Strategy Report on Research Infrastructures Roadmap 2018. Part 2: Landscape Analysis. Available at: http://roadmap2018.esfri.eu/media/1050/roadmap18-part2.pdf [Accessed 18 November 2021].

Goodson, H. Gödderz, K., Casino, A., Koureas, D. & Walton, S. Goodson (2020) Positioning DiSSCo among other research infrastructures. ICEDIG Deliverable D9.4. Available at: https://doi.org/10.5281/zenodo.3733033

Schulman, L., Lahti, K., Piirainen, E., Heikkinen, M., Ratio, O. & Juslén, A. (2021) The Finnish Biodiversity Information Facility as a best-practice model for biodiversity data infrastructures. *Scientific Data* **8**: 137(2021). https://doi.org/10.1038/s41597-021-00919-6

Smith, V. & Goodson, H. (2019) International landscape analysis of related international research infrastructures, supporting collections digitisation. ICEDIG Milestone MS46. Available at: https://icedig.eu/sites/default/files/milestone_ms46_icedig_international_landscape_analysis_of_related_international_research_infrastructures_.pdf [Accessed 18 November 2021].

APPENDIX 1

Appendix Table 1: Definitions of contact zones terms, which were sent to participating institutions.

Category	Term	Definition
General	Organisation	An entity – such as a company, an institution, or an association – comprising one or more people and having a particular purpose. In the context of this framework, this is the entity whose activity is being scored.
	Type	A high level class of information associated with a physical specimen held within a natural science collection.
	Phase	A stage with the data processing lifecycle.
	Infrastructure	The set of fundamental content, facilities, systems or services necessary for a community to function.
	Maturity Index	A measurement system used to assess the maturity level of a particular activity, domain or technology.
	Evidence	Examples relevant to the major 'Type' (not 'Phase') of activity, given as short unstructured text remarks and / or web links to further information.
Scope	Specimens	An evidential record of an individual, item, or part of a natural science collection.
	Collection registry/description (“Collections”)	Metadata used to describe any set of individuals, items, or parts (specimens) that form a whole or part of a natural science collection.
	Observations (opportunistic) (“Ad-hoc Observations”)	An evidential record of an unplanned encounter with an individual organism at a particular time and place.
	Observations (systematic) (“Formal Observations”)	An evidential record of an encounter with an individual organism at a particular time and place as part of a programme of study.
	Biological taxonomy/classification (“Taxonomy”)	Any activities associated with the a branch of science that encompasses the description, identification, nomenclature, and classification of organisms.

Category	Term	Definition
Scope	Biological descriptions/traits ("Traits")	The non-molecular phenotype of a biological entity, in the form of a text description, statement, multimedia or dataset.
	Geology	Any aspect of the characterization (including Earth or planetary system science) of rocks and minerals of any origin, in the form of a text description, statement or dataset.
	Molecular	Any aspect of the structure, function, evolution, mapping, and editing of an organism's DNA or RNA nucleotides.
	Literature	Any non fiction scholarly writing, or metadata associated with such writing, concerning any aspect of the natural world.
Phase	Create	The first stage in the data life cycle in which an initial digital representation is created.
	Aggregate	The bringing together of a group, body, or mass composed of many distinct parts or individuals.
	Access	The "ability to access" and benefit from some system or entity.
	Annotate	The addition of extra information associated with a particular point in any data, information or knowledge.
	Interlink	The connection of things (e.g entities in a database).
	Analyze	To subject to scientific analysis.
	Synthesis	The combining of often diverse conceptions into a coherent whole to create new knowledge.
Category	Data/Content	Factual information used as a basis for reasoning, discussion, or calculation.
	Standards	The rules (format and meaning) by which data are described, recorded and exchanged.

Category	Term	Definition
Category	Software	Any set of programs, procedures, and routines associated with the operation of a computer system.
	Hardware	Tools, machinery, and other durable equipment (e.g. computers and storage) associated with any phase of activity.
	Policy/Culture	The community networks and agreed practices to make our activities an openly shared, freely available, connected resource.
Maturity Index	P0 - No activity/inapplicable	No current/planned activity or inapplicable to an organisations operations.
	P1 - Planned	Named a strategy, roadmap or outline development as a proof of concept (evidenced through documentation or a prototype solution).
	P2 - Presence	Addresses part of the domain/problem set served, sometimes as a dependency to addressing other issues, and in use (evidenced through the use of the solution beyond the developing organisation).
	P3 - Performance	Addresses a majority/full scope of the domain it serves and in widespread use (evidenced through the richness of feature set and widespread use).
	P4 - Predominance	A domain leader to which all other innovators would aspire to or work with, addressing the full scope of the domain and sustained through continuous improvement (evidenced through market share).

SCORED BY: *<please populate>*

SCORES: P0 - No activity/inapplicable; P1 - Planned; P2 - Presence; P3 - Performance; P4 - Predominance

DEFINITIONS: For definitions of the sections, rows and columns, click or hover over the section, row or column header, or go directly to the 'Definitions' sheet.

FURTHER GUIDANCE: For more information, use this link to visit the FAQ document.

DATA / CONTENT

	Create		Aggregate		Access		Annotate		Interlink		Analyze		Synthesis		Evidence (i.e. links or examples to illustrate activity)
	Current	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*(Broad examples that illustrate your organisation's activity within the row's scope)*
Specimens	0	0	4	4	4	4	2	4	3	4	2	2	2	2	
Collection registry/description	0	0	2	4	2	4	2	4	2	4	2	3	2	2	
Observations (opportunistic)	0	0	4	4	4	4	2	4	3	4	2	2	2	2	
Observations (systematic)	0	0	4	4	4	4	2	4	3	4	2	2	2	2	
Biol. taxonomy/classification	0	0	3	4	3	4	2	4	2	4	2	3	0	0	
Biol. descriptions/traits	0	0	2	2	2	2	0	0	0	0	0	0	0	0	
Geology	0	0	2	2	0	0	0	0	2	2	0	0	0	0	
Molecular	0	0	2	2	2	2	0	0	0	0	0	0	0	0	
Literature	0	0	2	2	2	2	0	0	0	0	0	0	0	0	

STANDARDS

	Create		Aggregate		Access		Annotate		Interlink		Analyze		Synthesis		Evidence (i.e. links or examples to illustrate activity)
	Current	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*(Broad examples that illustrate your organisation's activity within the row's scope)*
Specimens	3	4	2	2	2	2	2	3	2	3	3	3	3	4	
Collection registry/description	2	3	2	3	2	2	2	2	3	4	2	3	2	4	
Observations (opportunistic)	3	4	2	2	2	2	2	3	2	3	3	3	3	4	
Observations (systematic)	3	4	2	2	2	2	2	3	2	3	3	3	3	4	
Biol. taxonomy/classification	3	4	2	2	2	2	2	4	2	4	3	4	3	4	
Biol. descriptions/traits	2	2	0	0	0	0	0	0	0	0	0	0	0	0	
Geology	2	2	0	0	0	0	0	0	0	0	0	0	0	0	
Molecular	2	2	0	0	0	0	0	0	0	0	0	0	0	0	
Literature	2	2	0	0	0	0	0	0	0	0	0	0	0	0	

SOFTWARE

	Create		Aggregate		Access		Annotate		Interlink		Analyze		Synthesis		Evidence (i.e. links or examples to illustrate activity)
	Current	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*Current*	*Ambition*	*(Broad examples that illustrate your organisation's activity within the row's scope)*
Specimens	4	4	2	2	3	3	2	2	2	3	2	2	2	2	
Collection registry/description	2	4	3	4	3	4	2	2	2	3	2	2	2	4	
Observations (opportunistic)	4	4	2	2	3	3	2	2	2	3	2	2	2	2	
Observations (systematic)	4	4	2	2	3	3	2	2	2	3	2	2	2	2	
Biol. taxonomy/classification	3	4	2	2	3	3	2	2	2	3	2	2	3	4	
Biol. descriptions/traits	2	2	0	0	0	0	0	0	0	0	0	0	0	0	

Appendix Figure 1: Screenshot showing a template of the survey which was sent to participating infrastructures. Each infrastructure rated their maturity index level for each phase, category and scope of activity, and were asked to provide evidence for this rating.