

DiSSCO

Distributed System of Scientific Collections



research infrastructure outline

February 2017

This document is intended for the purposes of early evaluation by national authorities, including governmental departments, research councils and academies.

Version
1702.4

Date
February 2017

For more information contact info@dissco.eu

Scientific rationale

Natural science collections are an integral part of the global natural and cultural capital. They include hundreds of millions of animals, plants, fossils, rocks, minerals and meteorites, which account for **55% of the natural sciences collections globally**, and represent **80% of the world bio- and geo-diversity**. Data derived from these collections underpin countless innovations, including tens of thousands of scholarly publications and official reports (used to support legislative and regulatory processes relating to health, food, security, sustainability and environmental change); inventions and products critical to our bio-economy; databases, maps and descriptions of scientific observations; instructional material for students, as well as educational material for the public.

In the last couple of decades, we have witnessed great changes in the way we work in biodiversity science. Remote sensing, rapid identification and molecular approaches allow us to efficiently monitor the changing world around us and understand the cause of those changes. Advances of digital, genomic and information technologies enable natural science collections to provide new discoveries and ask for new collection types and attributes while foster the development of new approaches to face the urgent societal challenges. As the volume and diversity of information deriving from natural science collections is exponentially increasing, so is the need for adequate infrastructures that go further than providing access to the different data classes. **A holistic approach is required, which will effectively underpin the entire research lifecycle and provide access to mass, linked, reliable and precise data.**

Our vision for DiSSCo is to unify European natural science collections, effectively transforming a dispersed and fragmented access model to an integrated data-driven pan-European research infrastructure.

A pan-European research infrastructure is urgently needed to solve the current limitations for the use of the data and related expertise contained by the natural science collections, largely caused by the still highly fragmented landscape in Europe and the very low level of digitisation. A pan-European research initiative, as impersonated by DiSSCo, enables faster delivery of information, widens the community of scientific users and broadens usage in general.

DiSSCo enables European natural science collections to operate as a versatile and integrated system of distributed nodes. The new research infrastructure **unifies access to information, provides new linked data associated with collections, and drives policy and process harmonisation**. As a result, the new RI achieves the economies of scope and scale necessary to maximise impact to science and society. Through digitisation, aggregation and linkage of European collections, critical new insights will enable scientists to address some of the world's greatest challenges.

Objectives and Services



DiSSCo in numbers

1.5 billion specimens

held in European collections

80% of global biodiversity

described based on European collections

100 collaborative projects

5,000 scientists

15,000 visiting scientists

3,000 scientific publications

10 million public visitors

annually

25 million web visitors

annually

The key high-level objectives of DiSSCo include:

- Bringing scientific collections to the information age, investing in a linked open data approach;
- Investing in balanced multi-modal access to collections;
- Improving researchers' capacity to use collection information to tackle complex scientific challenges;
- Supporting the interplay of social and cultural aspects of collection data;
- Developing and implementing targeted joint research agendas;
- Identifying collection data at European level and improving curation efficiency;
- Building and supporting paths to industrial innovation;
- Enhancing digital skills and competencies, tooling-up researchers to navigate the big data domain; and
- Engaging with society, providing alternative ways of benefiting from the national investments to collections.

DiSSCo will provide a set of four service classes available for a wide range of users: (i) Physical (transnational) and digital (virtual) access to collections; (ii) Joint research programming for data-driven scientific innovation; (iii) Training, support and engagement, and (iv) Policy harmonisation.

Service Deliverables

DiSSCo aims to support fluent and permanent collaboration and interoperability across all European natural science collections and to establish a model for unified collection access, which can be expanded to global scale, as other non-European regions promote in parallel infrastructures. Despite the homogeneous and

comprehensive accessibility to data, different user communities (Figure 1) have specific requirements that are to be met with six key services provided during the operational phase (Table 1).

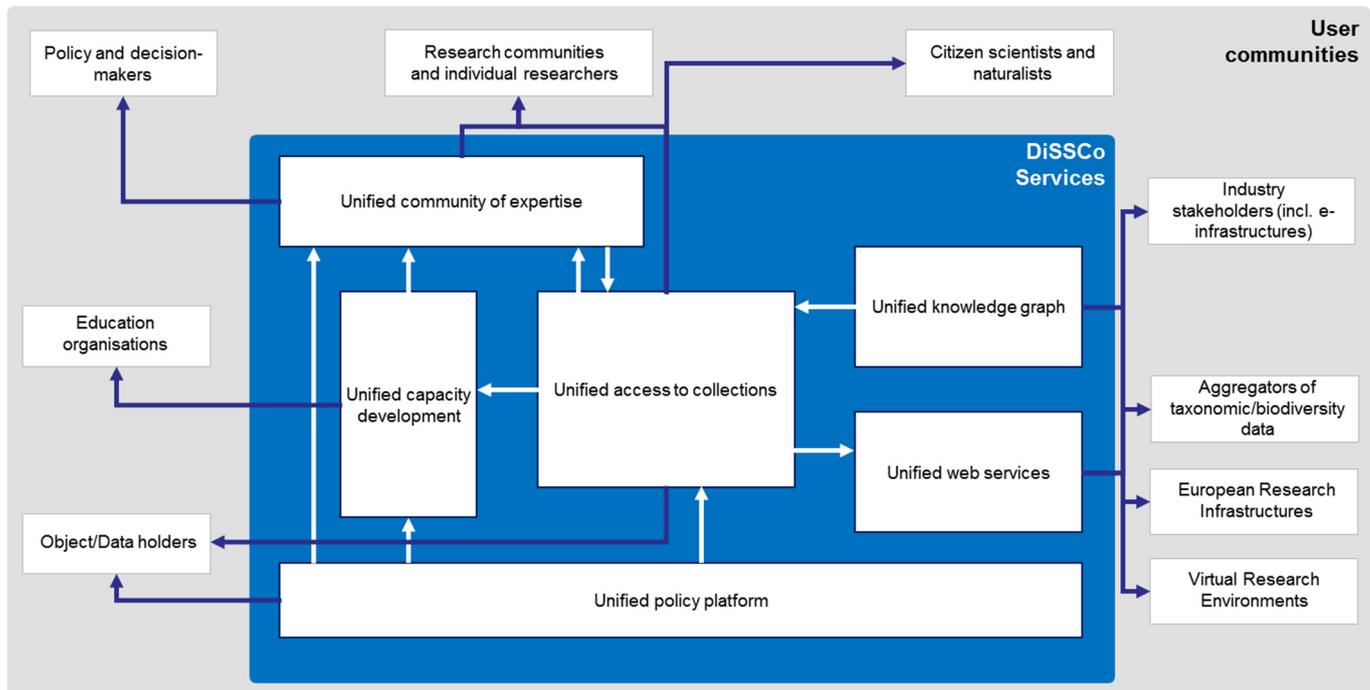


Figure 1. Interdependencies of service deliverables and links to identified user communities.

User communities

- **Research communities and individual researchers** in Environmental/natural sciences (incl. taxonomists, ecologists, bio-informaticians, conservationists, ethno-botanists, geneticists, chemists);
- **Virtual Research Environments (VREs)** users, providers and stakeholders;
- **Citizen scientists and naturalists** and coordinators of citizen science projects in biodiversity;
- **Object/Data holders** (partners to the project, external repositories, potential users of other domains)
- **Aggregators of taxonomic/biodiversity data** (e.g. GBIF, EoL) and Indexing agents (e.g. CoL, IPNI, ZooBank)
- **Policy and decision-makers** in governmental and non-governmental organisations
- **Education organisations** including vocational and academic teachers and students;
- **Industry stakeholders** including service providers, industrial devices/technology producers;
- **European Research Infrastructures** (incl. LifeWatch, ELIXIR, MIRRI, EPOS, and E-RIHS).

Table 1. Description of DiSSCo main service deliverables

Service deliverable	Short description
Unified policy platform	DiSSCo will be developed based on a common comprehensive policy corpus agreed to be implemented across all DiSSCo facilities. These policies will be underpinned by (i) open access principles, including FAIR principles, (ii) clear incentive schemes for citation and attribution, and (iii) thematic prioritisation, for the digitisation of the distributed collections. Furthermore, these policies will extend to the development of joint research programming activities across all its members.
Unified community of expertise	DiSSCo will facilitate collaborative research by enabling consistent pan-European digital access and data curation rights for research and curatorial and technical staff from all collections. DiSSCo will also create an integrated registry of experts with expertise linked to European collections.
Unified access to collections	Access to collections is a centrepiece service for DiSSCo. Services will be provided for both physical and virtual modes of access. DiSSCo will support a transnational access programme and couple this with comprehensive virtual (digital) access service. DiSSCo will deliver an information system which offers, to the fullest extent possible, consistent access for all authorised researchers not only to be able to access data from any collection, but also contribute to the curation and improvement of these data. A European Collections Data Portal will enable direct and open access to the integrated database.
Unified knowledge graph	The engineering of the DiSSCo knowledge graph will bring together digital information and assets from all participating institutions, with institutional management systems for collections, laboratories and other facilities remaining the persistent repository for these data, and will additionally include further digital knowledge derived from graphing these data objects and resolving identities between multiple institutions. The entire graph will be navigable as a set of interconnected rich data records.
Unified web services	The DiSSCo unified knowledge graph will be accessible via search, visualisation and download services to support efficient researcher access to data. DiSSCo will ensure that the resulting linked open data (LOD) is fully accessible for exploration, mining and semantic inference within European high-performance computing infrastructures, and fit for use by other European infrastructures (e.g. LifeWatch) and contributing to global infrastructures (e.g. GBIF).
Unified capacity development	A consistent pan-European approach to training will simplify accreditation of skills and enable appropriate mechanisms to be put in place for user access rights around DiSSCo data. The approach will also support the development of European collections personnel as a unified community of expertise and excellence.

Positioning

DiSSCo distributed facilities

The DiSSCo consortium is currently comprised of **62 natural science collection facilities** (natural science museums and botanical gardens) in **19 European countries** (Austria, Belgium, Bulgaria, Czech Republic, Germany, Denmark, Estonia, Spain, Finland, France, Greece, Italy, Netherlands, Norway, Poland, Portugal, Sweden, Slovakia, and United Kingdom). Other research collections-based institutions may join this common effort in the near future. These facilities participate in DiSSCo through the formulation of national nodes that have been set-up by means of MoUs or through the participation of pre-existing national-level consortia.



Access modes

Physical access

- **User visits** to collections and instruments;
- **User access to genetic samples** (tissue and DNA collections);
- **Loans**, sending specimens to the user for a certain time of use and under certain conditions.

Remote and virtual access

- **DiSSCo data portal**, containing geographical, morphological, genetic and chemical data, images, movies, sounds, links to other data sources;
- **Access to on-line registries and visualized collections**, allowing users to pre-study the collection parts of their interest;
- **On-demand digitisation**, including analysis of specimens.

Maturity and sustainability

DiSSCo is built on top of a strong, stable and cohesive European network of natural sciences collections-based institutions (CETAF) and benefits from the accumulated experience of several EC-funded projects such as SYNTHESYS I-III, EDIT, ViBRANT and pro-iBiosphere. **Together, this cumulative experience provides a robust corpus of technical, governance and socio-cultural feasibility reports and studies, that drives the**

development and operation of the new Research Infrastructure.

DiSSCo is uniquely placed among other distributed networks or infrastructures, as **the entirety of the participating facilities have secured long-term financial support through governmental statutory funding** and already successfully operate in the context of regional and/or national strategic priorities.

European Research Infrastructures: Landscape view

In the European landscape of environmental Research Infrastructures, different projects and landmarks describe services that aim at aggregating, monitoring, analysing and modelling geo and bio-diversity information. The effectiveness of these services, however, is based on the **quality and availability of primary reference data that today is scattered and incomplete**. DiSSCo provides the required taxonomic, bio-geographical and species trait data at the level of

precision and accuracy required to enable research for tackling grand societal challenges.

DiSSCo fills a significant gap in the landscape of the European research infrastructures. In the wider European RI landscape DiSSCo acts as the missing reliable block of the value chain where services and final products need to be grounded for adequately monitoring and modelling the geo- and bio-sphere. (Figure 2).

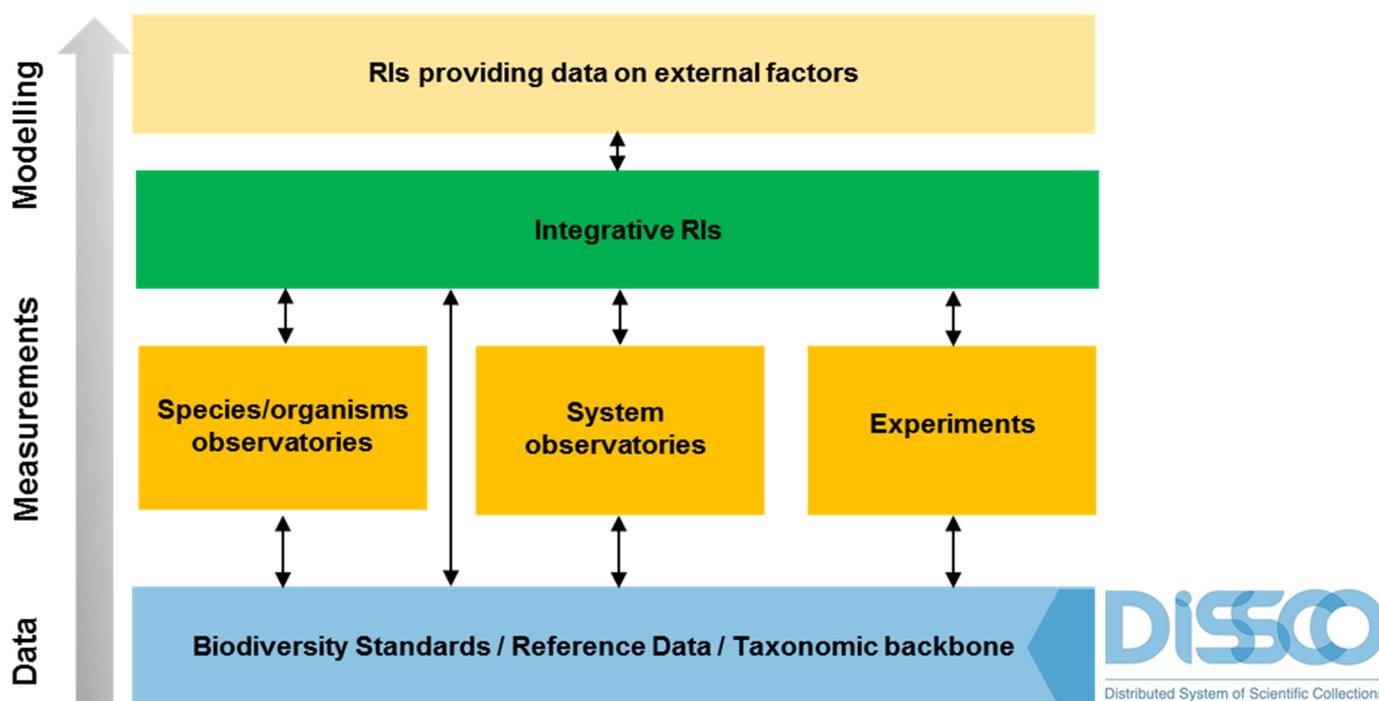


Figure 2. Simplified landscape view of the environmental research infrastructures. DiSSCo occupies the foundation layer, providing reference data and standards at the scale, precision and quality needed for other environmental monitoring and modelling infrastructures.

Implementation

The DiSSCo preparatory and construction phases are planned between 2018 and 2025 (Figure 3).

2018 - 2022

Preparatory phase (Innovation and Consolidation programmes):

Innovation programme: Refinement of technical design; deployment of pilot portal and service test beds; update of business plan; contact-building with industrial stakeholders; technical innovation for mass scale digitisation, automation, robotics, and data models.

Consolidation programme: Refinement of governance structure; preparation for setting-up of legal entity; communication and capacity building; update of business plan; harmonisation of nodes policies and processes (incl. access and training); site selection and staff recruitment; development of national-level collection and investment strategies.

2019 – 2024

Construction phase: Implementation of national/regional investment plans for infrastructure upgrades and large scale digitisation programmes; application of joint DiSSCo programmes and policies, quality control and risk management; establishment of regional/thematic hubs; active membership; construction of the DiSSCo Hub (including all services).

2024 – 2025

Deployment phase: Transition from preparatory phase to the fully operational governance structures; modest/soft start of operations.

2025 onwards

Operational phase: Full deployment of operations and services; first full evaluation process; review of organisation and business model; full scale operations.

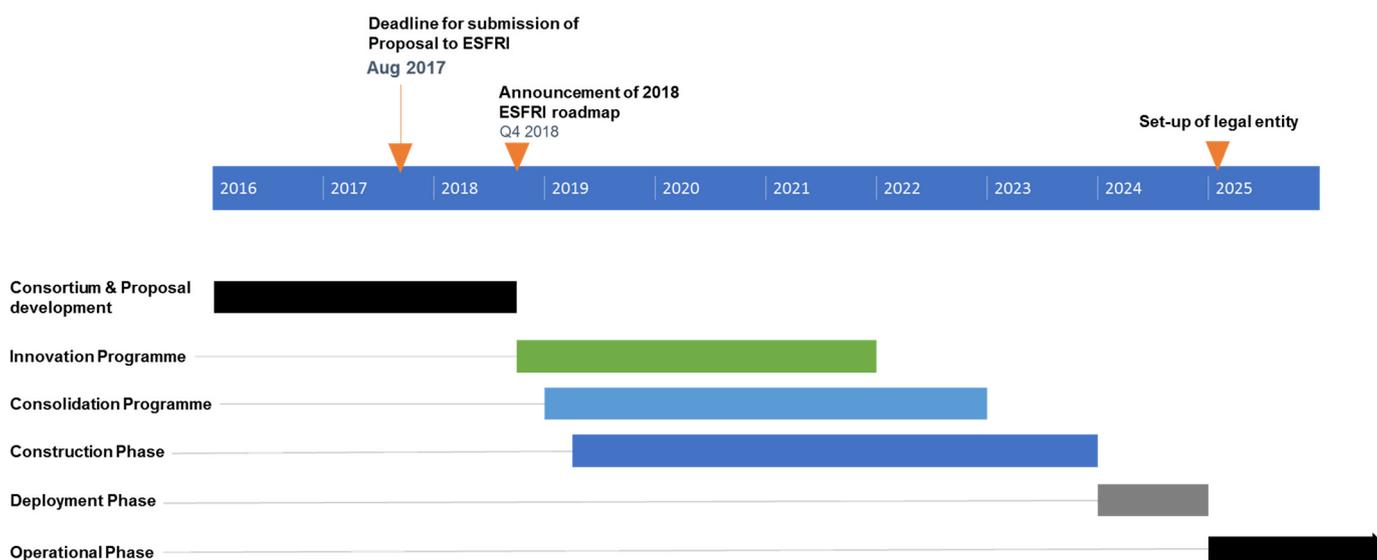


Figure 3. Timeline of the DiSSCo different phases from preparatory to operational, organised in distinct and interlinked programmes (Innovation, Consolidation and Construction).

Governance model and research infrastructure topology

The DiSSCo governance model will aim to:

- Achieve the maximum possible inclusion of and balance between member states and associated countries in the decision-making processes;
- Ensure transparent and fair procedures, with clearly indicated accountable parties;
- Minimise the administrative overheads, without compromising the integrity of the governing procedures;
- Maintain adequate flexibility in order to adapt to future change of needs, including fast expansion of membership or widening of scope, and;
- Set up well-defined consensus reaching and conflict resolution mechanisms.

A **central shared coordination approach** will be followed whereby different distributed nodes are coordinated by a separate organisation. The topology of the RI is based on a hub and spoke model (**Figure 4**).

A **central facility/headquarters** will be responsible for the **coordination of all node-related activities** while resources will be allocated separately to both the central coordination office (CCO) and the partners, in support of the implementation phase and operational aspects of the infrastructure.

The application of this model will be predicated on existence of a robust governance structure. Clear governance and management practices are jointly adopted by all participating facilities through a series of Memoranda of Understanding (MoU). These MoUs also highlight the common approach in formalising the bodies to govern the planned DiSSCo legal entity.

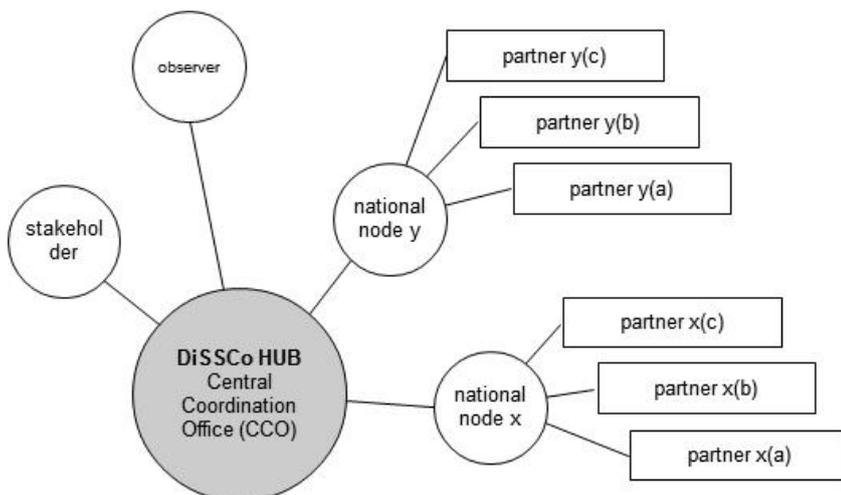


Figure 4. Topology of DiSSCo RI, with a hub (Central Coordination Office) and a series of spokes (national nodes/partner facilities).

Polygrapha
cyanca

Im. et Salv.
Hualla



Vanessa
myrrina

Doubleday



gyra

astillia
myia



Info@disco.eu | <http://disco.eu> | @DiSSCoEU