I Jornada de Ciència Oberta a la UIB **Impulsant la recerca responsable, transparent i col·laborativa**

Palma de Mallorca, November 24th, 2023

Towards digitally unifying European natural history collections

Rafael Zardoya
Director

Open science *is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process.*

Michael Nielsen, 2011

OECD
BETTER POLICIES FOR BETTER LIVES

1/ Data governance for trust

To develop and implement coordinated mechanisms, strategies, or policies to make research data* openly accessible and reusable

2/ Technical standards and practices

To promote and require compliance with technical standards and practices that make research data* findable, accessible, interoperable and re-usable (FAIR)

3/ Incentives and rewards

To foster and support the development and implementation of effective models of reward and recognition that provide incentives to grant access to research data*

4/ Responsibility, ownership, and stewardship

To allocate responsibility, ownership, and stewardship, while also tailoring and implementing licensing to protect research data* producers' rights.

5/ Sustainable infrastructures

To develop strategies, road-maps, funding plans, and business models, to ensure sustainable infrastructures for research data*, including data and software repositories and services, free of charge at the point of use.

6/ Human capital

To support the development of the human capital necessary to realize the full potential benefits of enhancing access to research data*

7/ International cooperation

To collaborate at the international level on access to research data*, notably where making use of datasets across borders can help the advancement of science and contribute to solving global societal challenges.

* research data and other research-relevant digital objects from public funding

ESFRI

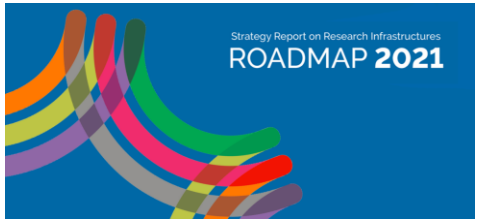
European Strategy Forum on Research Infrastructures

- A strategic instrument to develop the **scientific integration** of Europe and to strengthen its **international outreach**.
- Established in 2002, with a mandate from the EU Council to support a coherent and strategy-led approach to **policy-making on research infrastructures** (RI), and to facilitate multilateral initiatives leading to the better use and development of RI
- The competitive and **open access to high quality RI** supports and benchmarks the quality of the activities of European scientists, and attracts the best researchers from around the world.

ESFRI

European Strategy Forum on Research Infrastructures

- A strategic instrument to develop the **scientific integration** of Europe and to strengthen its **international outreach**.
- Established in 2002, with a mandate from the EU Council to support a coherent and strategy-led approach to **policy-making on research infrastructures** (RI), and to facilitate multilateral initiatives leading to the better use and development of RI
- The competitive and **open access to high quality RI** supports and benchmarks the quality of the activities of European scientists, and attracts the best researchers from around the world.



The Strategy Report on Research Infrastructures includes the **Roadmap** with ESFRI Projects and ESFRI Landmarks and the ESFRI vision of the **evolution** of Research Infrastructures in Europe, addressing the mandates of the European Council, and identifying **strategy goals**.

Part 1
STRATEGY REPORT

Part 2
LANDSCAPE ANALYSIS

Part 3
PROJECTS & LANDMARKS

PART3

PROJECTS & LANDMARKS

DATA, COMPUTING & DIGITAL RESEARCH INFRASTRUCTURES

169 EBRAINS
170 SLICES
171 SoBigData++

ENERGY

172 IFMIF-DONES
173 MARINERG-i

ENVIRONMENT

174 DANUBIUS-RI
175 DiSSCo
176 eLTER RI

2018

HEALTH & FOOD

177 EIRENE RI
178 EMPHASIS
179 EU-IBISBA
180 METROFOOD-RI

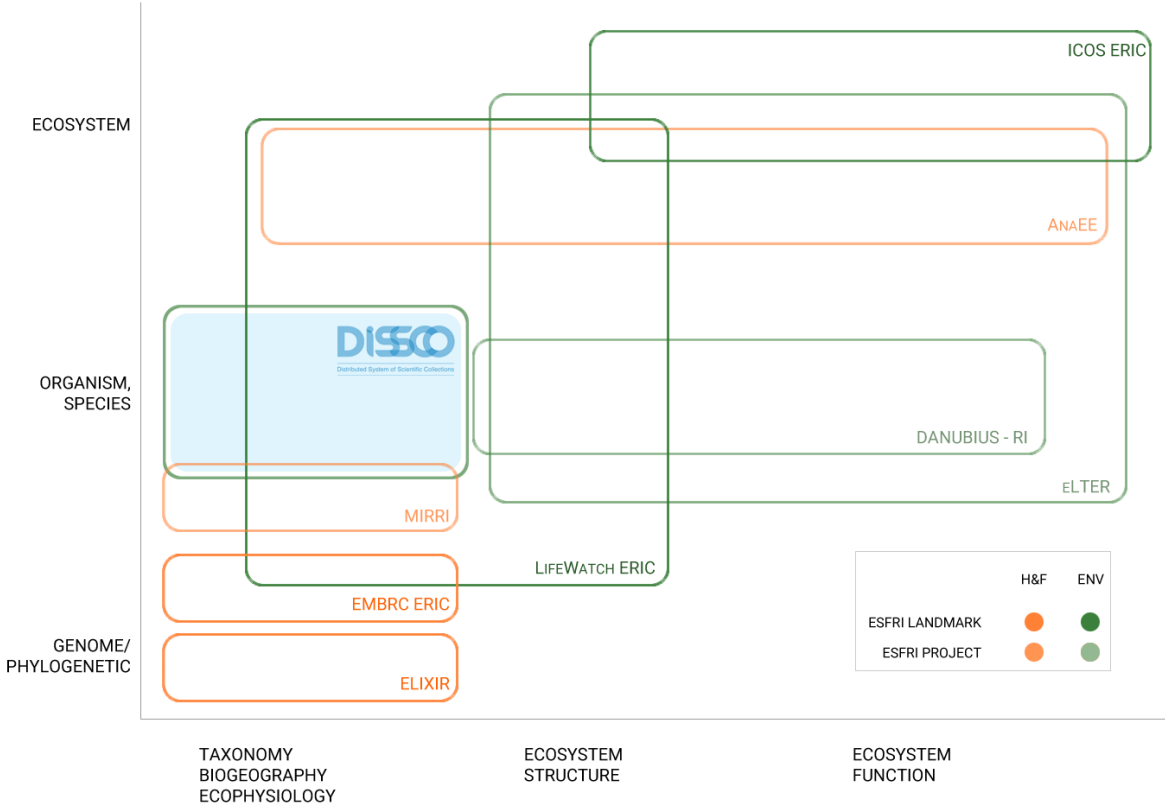
PHYSICAL SCIENCES & ENGINEERING

181 EST
182 ET
183 EuPRAXIA
184 KM3NeT 2.0

SOCIAL & CULTURAL INNOVATION

185 E-RIHS
186 EHRI
187 GGP
188 GUIDE
189 OPERAS
190 RESILIENCE

DiSSCo's place among other Research Infrastructures



GBIF Global Biodiversity Information Facility

GBIF

GBIF—the Global Biodiversity Information Facility—is an international network and Research Infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.

iDigBio Integrated Digitized Biocollections

iDigBio

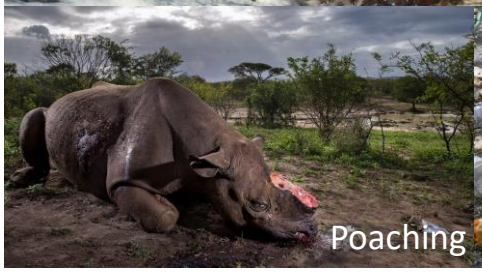
iDigBio – through a dedicated programme – makes data and images for millions of biological specimens available. The mission of iDigBio is furthermore to develop a national infrastructure by overseeing implementation of standards and best practices for digitisation.

Catalogue of Life

Catalogue of Life (COL)

Catalogue of Life (COL) is a collaboration bringing together the effort and contributions of taxonomists and informaticians from around the world.

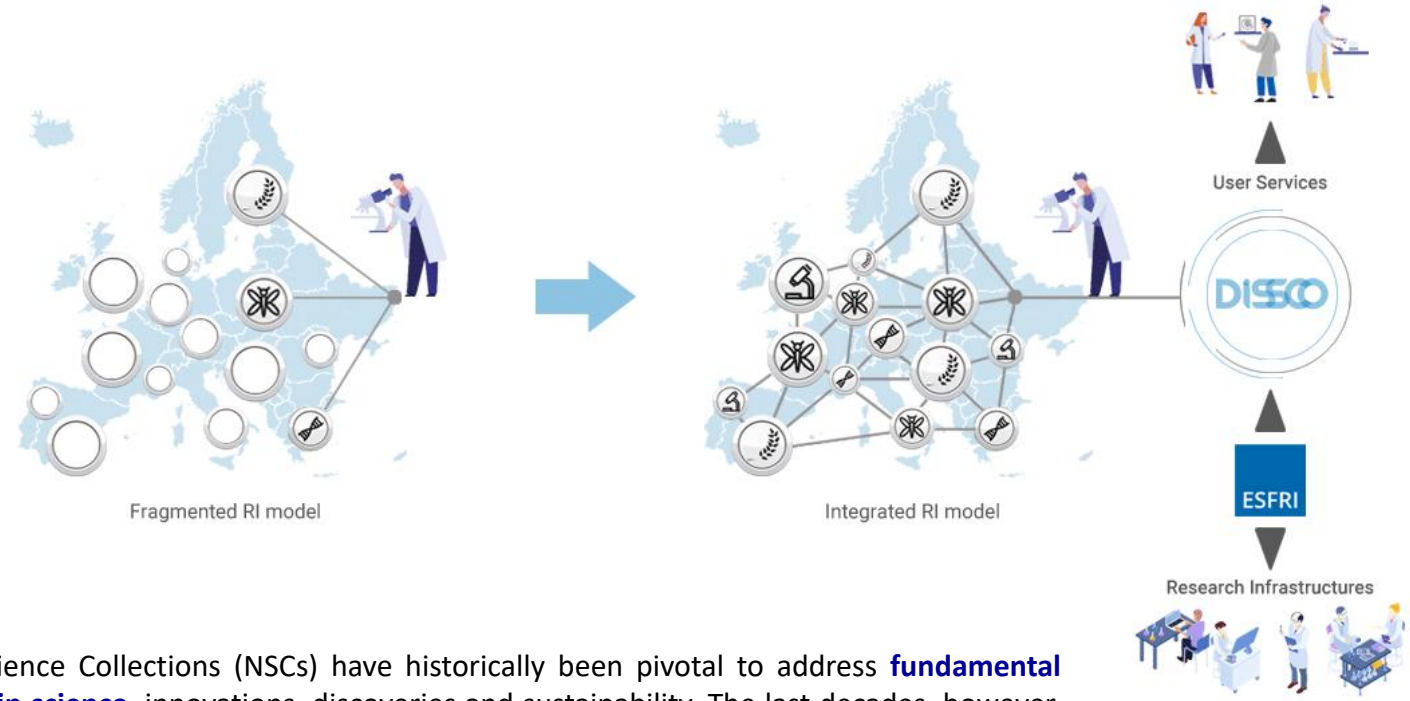
The environmental crisis



SUSTAINABLE DEVELOPMENT GOALS

The 17 Sustainable Development Goals (SDGs) were adopted by the United Nations in 2015 as a universal call to action to end poverty, **protect the planet**, and ensure that by 2030 all people enjoy peace and prosperity. They recognize that action in one area will affect outcomes in others, and that development must balance social, economic and **environmental sustainability**.

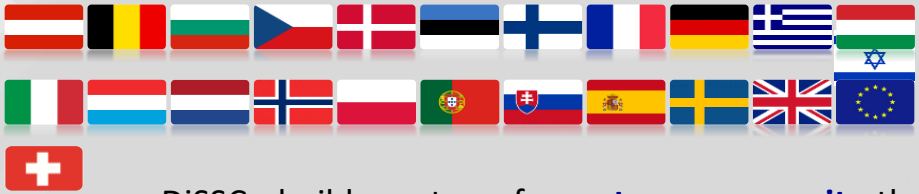
Towards a new model of research in natural sciences



Natural Science Collections (NSCs) have historically been pivotal to address **fundamental questions in science**, innovations, discoveries and sustainability. The last decades, however, have witnessed a **dramatic change in research practice** that calls for a novel, integrative approach where **cross-linked information** effectively underpins the entire research life cycle and provides **open access** to mass and precise data.

DiSSCo

Distributed System of Scientific Collections



DiSSCo builds on top of a **mature community** that seeks for **strategic collaboration** giving value to the **open access to geo and biodiversity assets and data** by ordering a common knowledge base. It represents the largest ever formal agreement between natural science collection facilities

Objectives

- Create a one-stop e-science infrastructure providing discovery, access, interpretation, and analysis of complex linked data.
- Provide **end-user services** such as digitization on demand, research support and training activities to address current community limitations.
- Optimize **collection access, curation and management** practices in individual institutions, enabling strategies under a common research agenda.
- Accelerate **digitization,** taking the current workflows to an industrial scale.
- Permanently link representations of **digital specimens** to their attributes across distributed digital resources, thus ensuring robust science.
- Reduce the global **carbon footprint** with digital collections access that will reduce international trips and global shipments of specimens.
- Improve efficiency, facilitate economies of scale, make natural science research more responsive and resilient to urgent needs and accelerate **biodiversity discovery**.

Combining historic collection data with data from new technologies

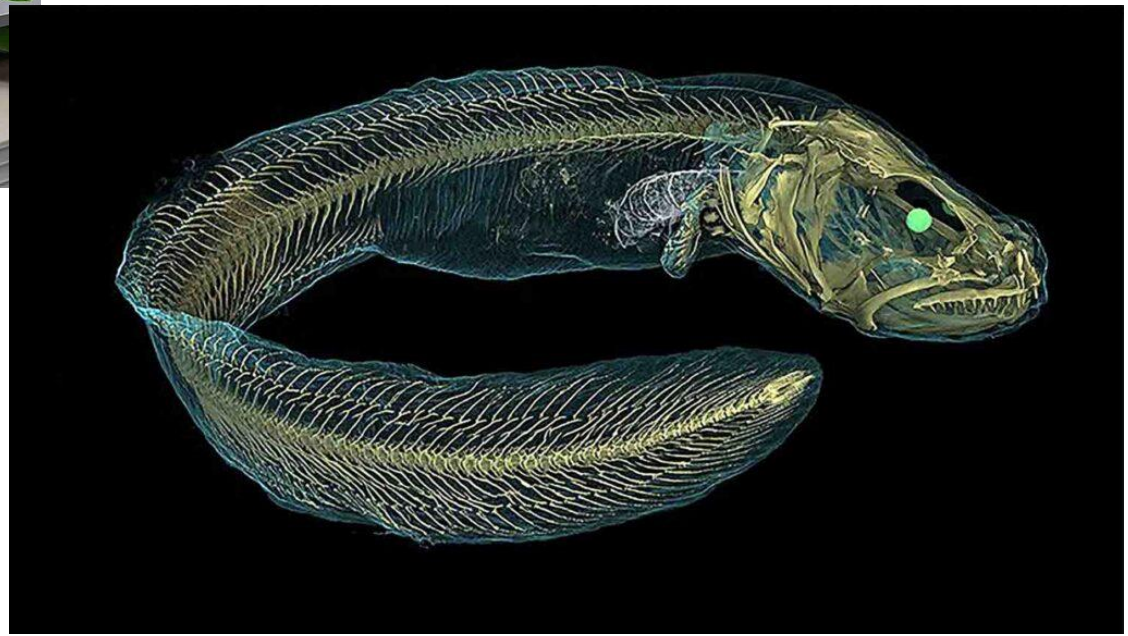
+

- Genomic data
- Biochemical data
- Morphological data
- Geographical data
- Taxonomic information
- Species interactions data
- Ecological data

→

- Accelerated field species identification, regular environmental monitoring, trend analysis and future prediction.
- Machine readability will enable integration of quality assured FAIR data into analytical workflows and tools.

Digital twin



Requirements

BRINGING NSCs TO THE INFORMATION AGE

1.5bn objects to be digitised.

90 petabytes of new data.

5-15K unique users every day.

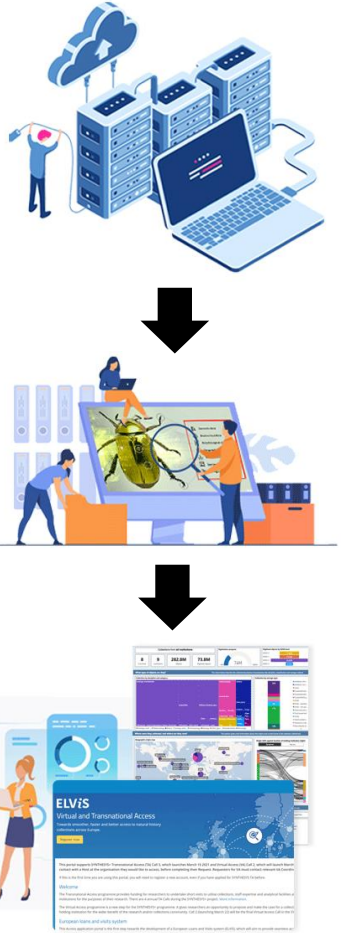
=

NEED FOR A ROBUST TECHNICAL INFRASTRUCTURE

To support working with digital specimens and collections over their entire research data life cycle.

To provide unified open access to the digital information, ensuring that it is FAIR.

Technical infrastructure



- Repositories with data provided by participating institutions

DiSSco will **connect data** provided by its participating institutions in trusted repositories. These can include local institutional repositories as well as global thematic repositories such as GBIF. It will also connect data in third-party repositories like genetic sequence and literature databases. All data that can be linked to collection objects (specimens) are in scope.

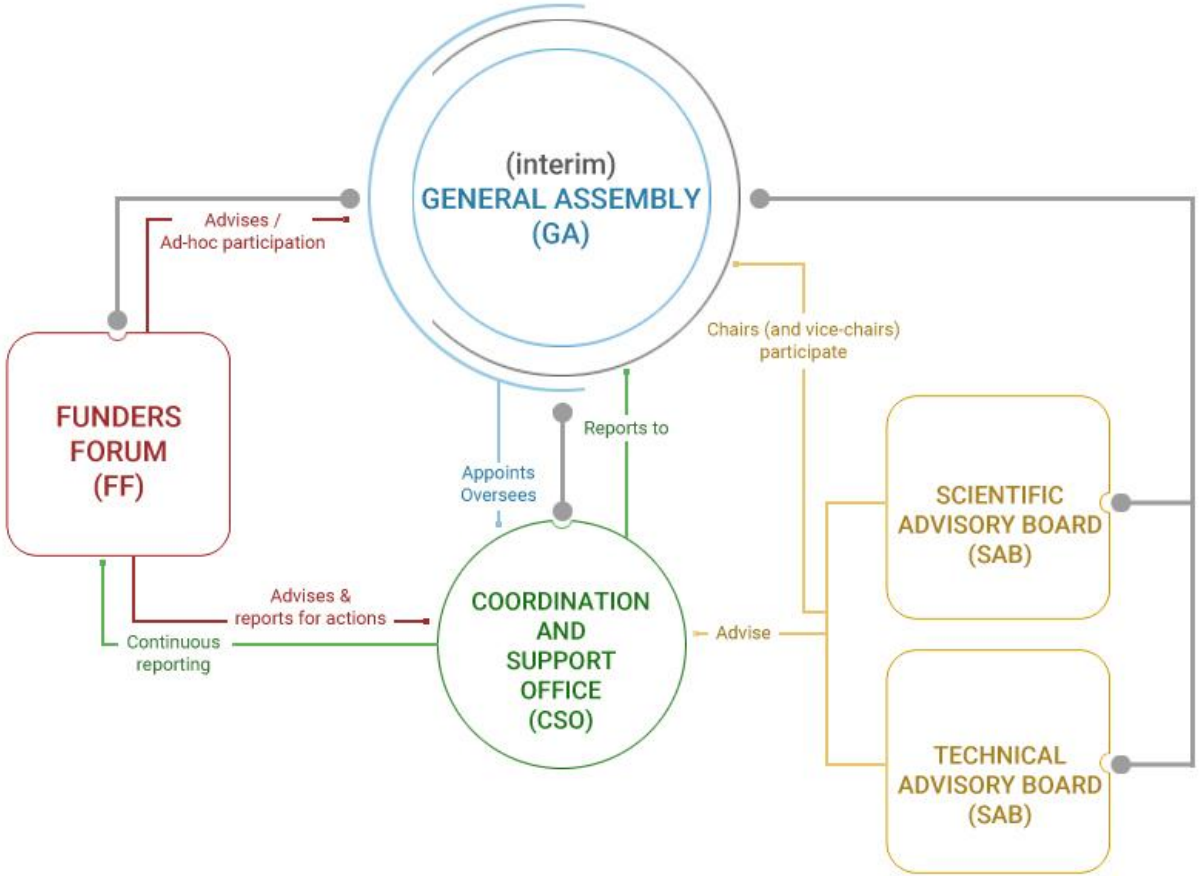
- Digital Object Infrastructure

The data will be linked through a **Digital Object (DO) infrastructure** in which Digital Collection (DC) objects and Digital Specimen (DS) objects are the principal object types, and the basis for a Natural Sciences Identifier Registry (NSIDR). The DO infrastructure will include tools for federation and linkage as well as services to support annotation and enrichment of the data by the scientific community.

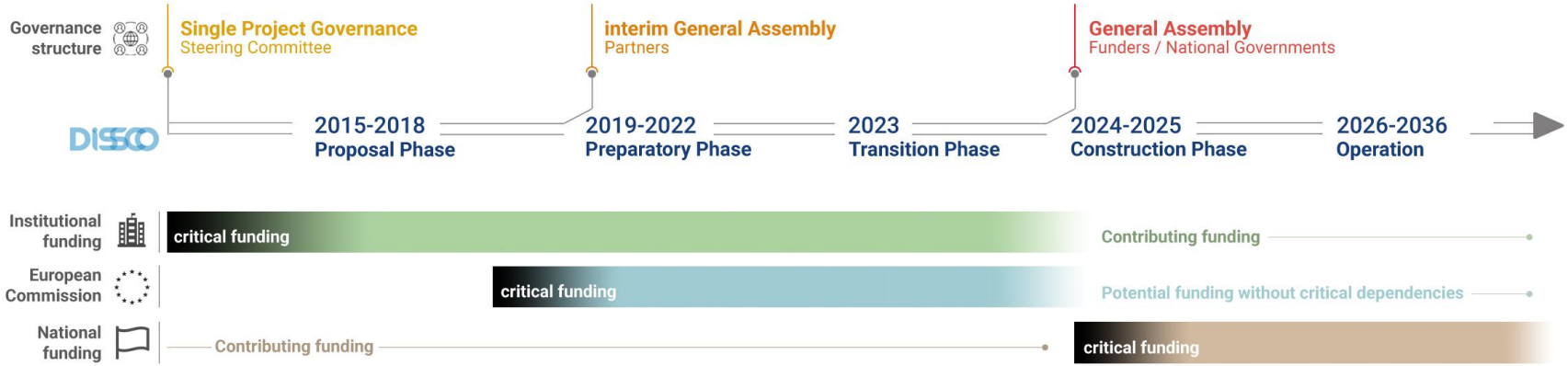
- Community Services

The infrastructure will provide **community services to discover, consume and interact** with the federated Digital Collection and Digital Specimen data. NSCs providers will become an integral part of the European scientific community, and better equipped for changing user needs and new scientific usage of collections.

DiSSCo's governance



DiSSCo's roadmap



DiSSCo ES
Spanish Distributed System of Scientific Collections

DiSSCo-ES Miembros Comunicación Coordinación Recursos

Who we are?

CSIC

Universidad de Navarra

UNIVERSIDAD COMPLUTENSE MADRID

VNIVERSITAT DE VALÈNCIA

museu de ciències naturals de Barcelona

Cabildo de Gran Canaria

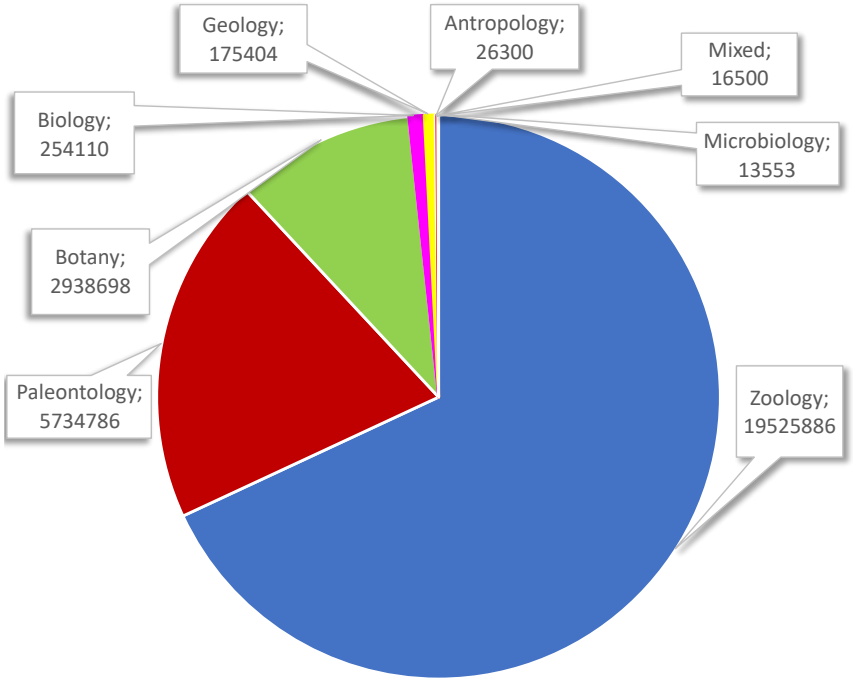
6 organizations and 27 facilities

DiSSCo ES

Spanish Distributed System of Scientific Collections

28,685,237

objects

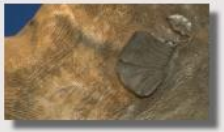


MNCN collections (>11 million)

Mammals

Birds

Reptiles


Fishes

Insects

Mollusks

Arthropods

Invertebrates

Nematodes

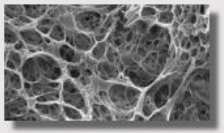
Fossil vertebrates

Fossil Invertebrates

Paleobotany

Prehistory

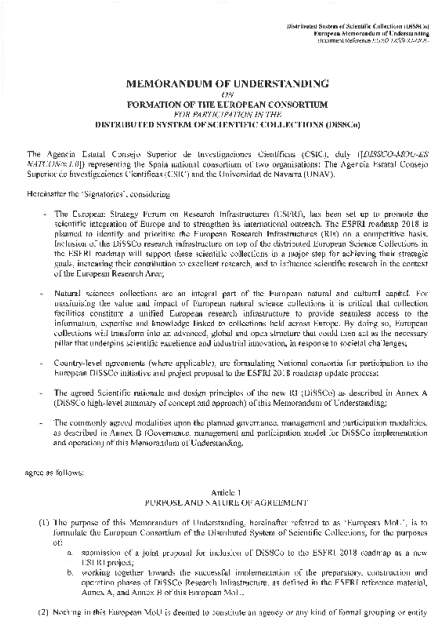
Geology


Biobank

Sound library

DiSSCo ES

Spanish Distributed System of Scientific Collections



Distributed System of Scientific Collection (DiSSCo)
European Memorandum of Understanding
Document reference *[NAD-MOU-DOC]*

MEMORANDUM OF UNDERSTANDING
ON
FORMATION OF THE EUROPEAN CONSORTIUM
FOR PARTICIPATION IN THE
DISTRIBUTED SYSTEM OF SCIENTIFIC COLLECTIONS (DISSCo)

The Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC), duly *[DISSCO-MOU-ES NATIONAL]* representing the Spain national consortium of two organisations: The Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC) and the Universidad de Navarra (UNAV).

Hereinafter the 'Signatories', considering

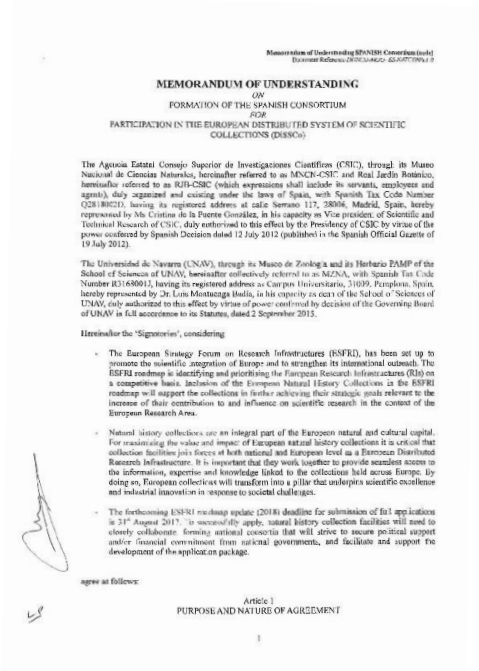
- The European Strategy Forum on Research Infrastructures (ESFRI), has been set up to promote the scientific integration of Europe and to strengthen its international outreach. The ESFRI roadmap 2018 is planned to identify and prioritise the European Research Infrastructures (RIs) on a competitive basis. Inclusion of the DiSSCo research infrastructure on top of the distributed European Science Collections in the ESFRI roadmap will support these scientific collections in a major step for achieving their strategic goals, increasing their contribution to excellent research, and to influence scientific research in the context of the European Research Area;
- Natural sciences collections are an integral part of the European natural and cultural capital. For maximising the value and impact of European natural science collections it is critical that collection facilities constitute a unified European research infrastructure to provide seamless access to the information, expertise and knowledge linked to collections held across Europe. By doing so, European collections will transform into an advanced, global and open structure that could then act as the necessary pillar that underpins scientific excellence and industrial innovation, in response to societal challenges;
- Country-level agreements (where applicable), are formulating National consortia for participation to the European DiSSCo initiative and project proposal to the ESFRI 2018 roadmap update process;
- The agreed Scientific rationale and design principles of the new RI (DiSSCo) as described in Annex A (DiSSCo high-level summary of concept and approach) of this Memorandum of Understanding;
- The commonly agreed modalities upon the planned governance, management and participation modalities, as described in Annex B (Governance, management and participation model for DiSSCo implementation and operation) of this Memorandum of Understanding,

agree as follows:

Article 1
PURPOSE AND NATURE OF AGREEMENT

(1) The purpose of this Memorandum of Understanding, hereinafter referred to as 'European MoU', is to formulate the European Consortium of the Distributed System of Scientific Collections, for the purpose of:
   a. submission of a joint proposal for inclusion of DiSSCo to the ESFRI 2018 roadmap as a new ESFRI project;
   b. working together towards the successful implementation of the preparatory, construction and operation phases of DiSSCo Research Infrastructure, as defined in the ESFRI reference material, Annex A, and Annex B of this European MoU.

(2) Nothing in this European MoU is deemed to constitute an agency or any kind of formal grouping or entity



Memorandum of Understanding SPANISH Consortium (node)
Document Reference *[DISSCO-MOU-ES-SIGNATORY]*

MEMORANDUM OF UNDERSTANDING
ON
FORMATION OF THE SPANISH CONSORTIUM
FOR
PARTICIPATION IN THE EUROPEAN DISTRIBUTED SYSTEM OF SCIENTIFIC COLLECTIONS (DiSSCo)

The Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC), through its Museo Nacional de Ciencias Naturales, hereinafter referred to as MNCN-CSIC and Real Jardín Botánico, hereinafter referred to as RJB-CSIC (which expressions shall include its servants, employees and agents), duly organized and existing under the laws of Spain, with Spanish Tax Code Number Q2818002D, having its registered address at calle Serrano 117, 28006, Madrid, Spain, hereby represented by Ms Cristina de la Puente González, in his capacity as Vice president of Scientific and Technical Research of CSIC, duly authorised to this effect by the Presidency of CSIC by virtue of the power conferred by Spanish Decision dated 12 July 2012 (published in the Spanish Official Gazette of 19 July 2012).

The Universidad de Navarra (UNAV), through its Museo de Zoología and its Herbario PAMP of the School of Sciences of UNAV, hereinafter collectively referred to as MZNA, with Spanish Tax Code Number R3160001J, having its registered address as Campus Universitario, 31009, Pamplona, Spain, hereby represented by Dr. Luis Monteagudo Badía, in his capacity as dean of the School of Sciences of UNAV, duly authorized to this effect by virtue of power conferred by decision of the Governing Board of UNAV in full accordance to its Statutes, dated 2 September 2015.

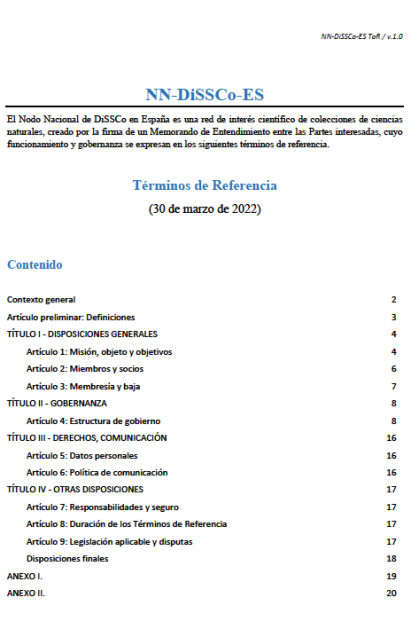
Hereinafter the 'Signatories', considering

- The European Strategy Forum on Research Infrastructures (ESFRI), has been set up to promote the scientific integration of Europe and to strengthen its international outreach. The ESFRI roadmap is identifying and prioritising the European Research Infrastructures (RIs) on a competitive basis. Inclusion of the European Natural History Collections in the ESFRI roadmap will support the collections in further achieving their strategic goals relevant to the increase of their contribution to and influence on scientific research in the context of the European Research Area.
- Natural history collections are an integral part of the European natural and cultural capital. For maximising the value and impact of European natural history collections it is critical that collection facilities join forces at both national and European level as a European Distributed Research Infrastructure. It is important that they work together to provide seamless access to the information, expertise and knowledge linked to the collections held across Europe. By doing so, European collections will transform into a pillar that underpins scientific excellence and industrial innovation in response to societal challenges.
- The forthcoming ESFRI roadmap update (2018) deadline for submission of full applications is 21st August 2017. To successfully apply, natural history collection facilities will need to closely collaborate, forming national consortia that will strive to secure political support and/or financial commitment from national governments, and facilitate and support the development of the application package.

agree as follows:

Article 1
PURPOSE AND NATURE OF AGREEMENT

1



NN-DiSSCo-ES ToR / v.1.0

NN-DiSSCo-ES

El Nodo Nacional de DiSSCo en España es una red de interés científico de colecciones de ciencias naturales, creado por la firma de un Memorando de Entendimiento entre las Partes interesadas, cuyo funcionamiento y gobernanza se expresan en los siguientes términos de referencia.

Términos de Referencia

(30 de marzo de 2022)

Contenido

Contexto general	2
Artículo preliminar: Definiciones	3
TÍTULO I - DISPOSICIONES GENERALES	4
Artículo 1: Misión, objeto y objetivos	4
Artículo 2: Miembros y socios	6
Artículo 3: Membresía y baja	7
TÍTULO II - GOBERNANZA	8
Artículo 4: Estructura de gobierno	8
TÍTULO III - DERECHOS, COMUNICACIÓN	16
Artículo 5: Datos personales	16
Artículo 6: Política de comunicación	16
TÍTULO IV - OTRAS DISPOSICIONES	17
Artículo 7: Responsabilidades y seguro	17
Artículo 8: Duración de los Términos de Referencia	17
Artículo 9: Legislación aplicable y disputas	17
Disposiciones finales	18
ANEXO I.	19
ANEXO II.	20

DiSSCo

DiSSCo ES

ToR

Transdisciplinary research

ARQ/Archeology

BDV / Biodiversity
CTA / Earth and water sciences
MAR / Marine science and technologies
POL / Polar Research

ALI / Food Science and Technology
AYF / Agriculture and forestry
GYA / Livestock and aquaculture

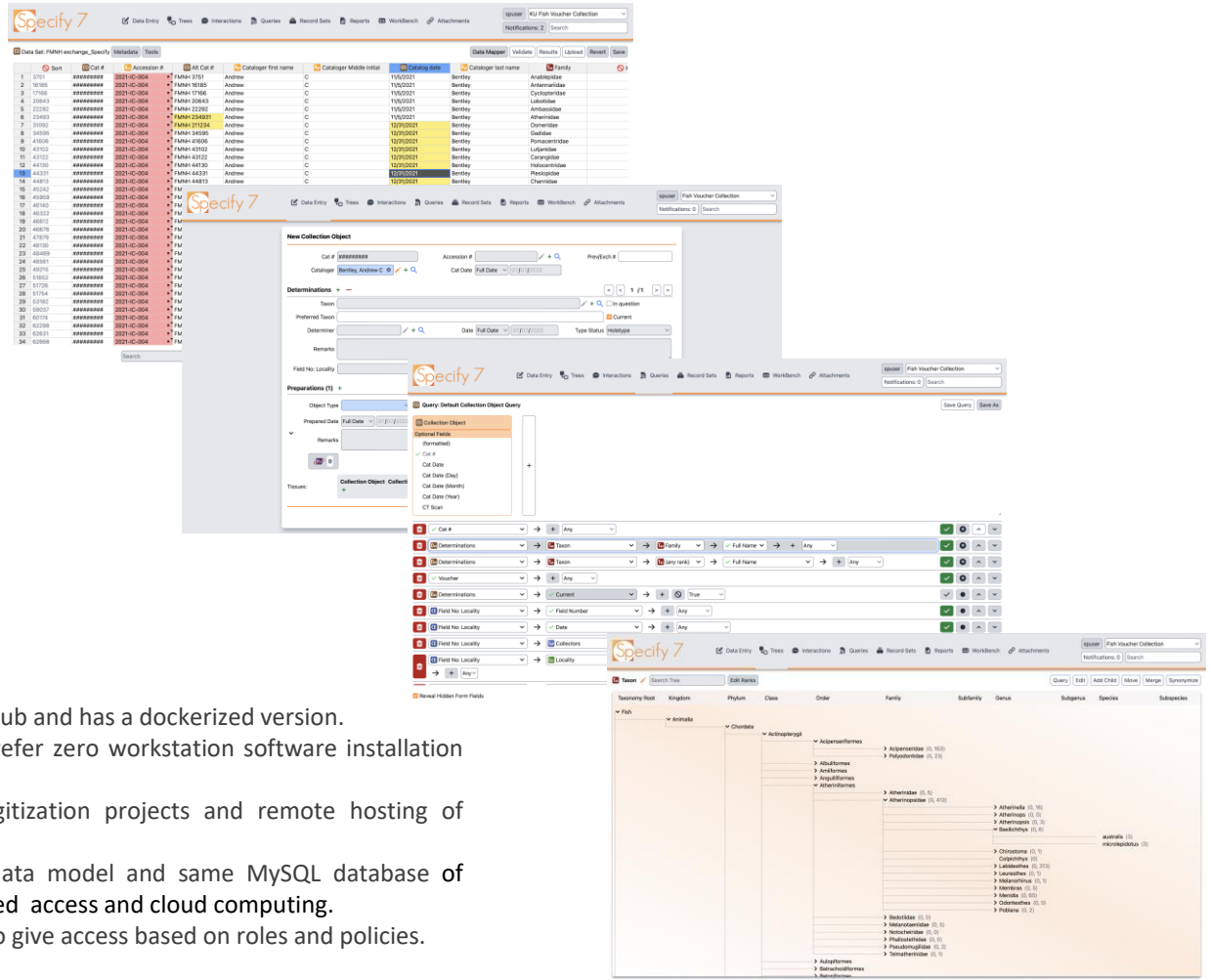
BIF / Integrative Biology and Physiology
BMC / Molecular and Cell Biology
BTC / Biotechnology

FOS / Physiopathology of organs and systems
IIT / Immunity, infection and immunotherapy

Specify COLLECTIONS CONSORTIUM

Founding Partners

- Commonwealth Scientific and Industrial Research Organisation
- University of Florida
- Consejo Superior de Investigaciones Científicas
- University of Kansas
- Danske Naturhistoriske Museer
- University of Michigan
- Muséum d'histoire naturelle de Genève



- The server code is open source and available from GitHub and has a dockerized version.
- It is a great starting platform for collections which prefer zero workstation software installation and ubiquitous web browser access.
- The browser architecture supports collaborative digitization projects and remote hosting of institutional or project specimen databases.
- It uses the same interface layout language, same data model and same MySQL database of previous versions but adds the ease-of-use of web-based access and cloud computing.
- The Security and Accounts tool allows administrators to give access based on roles and policies.
- It meets international accessibility standards.

https://www.youtube.com/watch?v=FpuXxZO-8oc