



DISCO

Distributed System of Scientific Collections



INFORMATION DAY

Distributed System of Scientific Collections



Vienna, AT, 5 December 2019



Natural Science Collections support **discovery** and **modelling** of all life on earth



DiSSCo provides access to biodiversity data at the scale and precision required

That's what counts

DiSSCo by the Numbers

1,500,000,000

specimens

119

participating
institutions

5,000

full-time scientists

21

countries

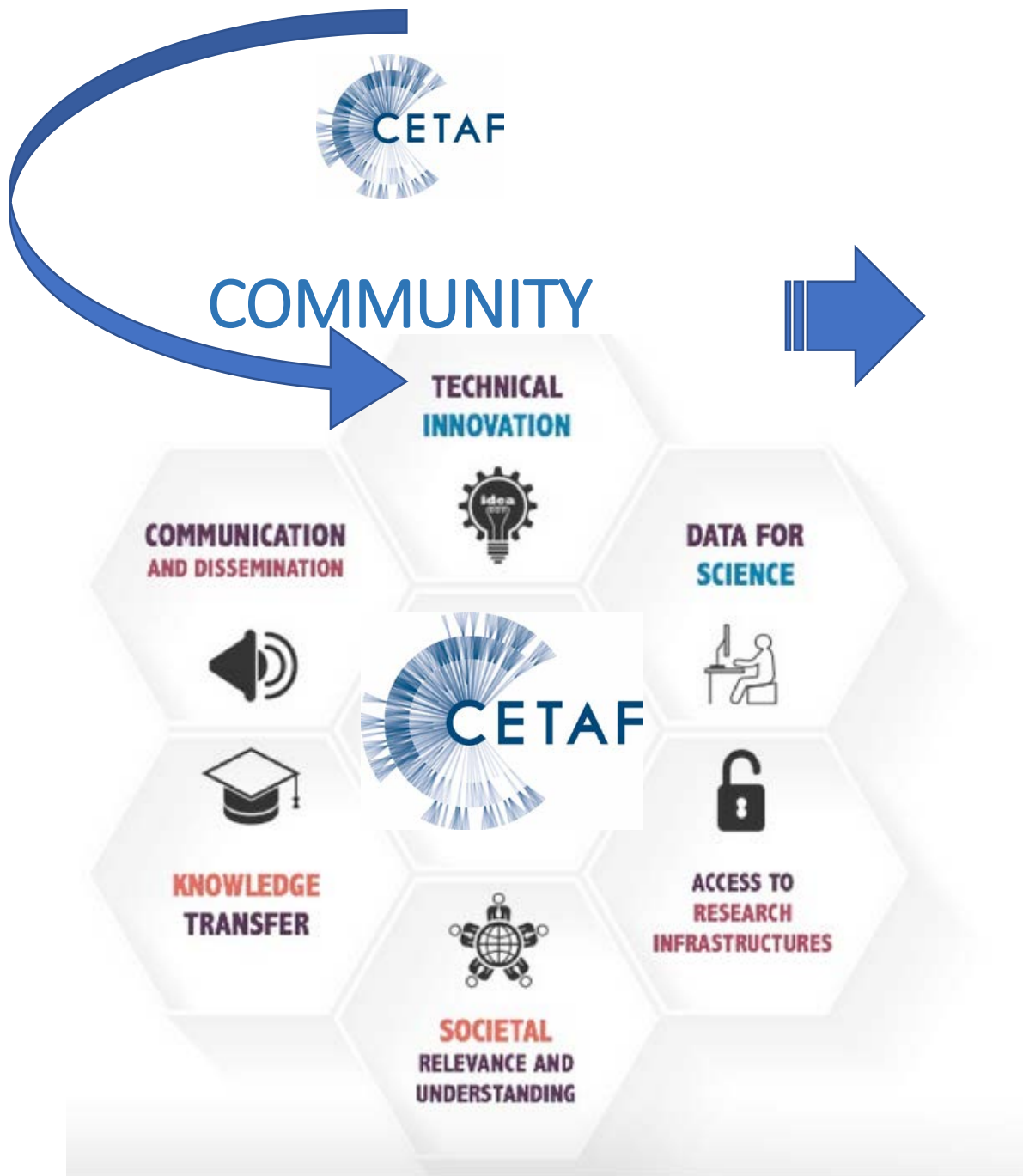
- > 100 collaborative projects
- > 6,000 scientific visitors pa
- > 3,000 publications pa
- > 10 million public visitors pa
- > 25 million web visitors pa



Data containers
Reliable reference
Irreplaceable uniqueness
For multiple and cross-cutting issues



Successful initiatives relies on the strong **community** behind and their **collaborative work**



INITIATIVES - PROJECTS



Preparing RI through a complex programme
(EUR 20 million)



DiSSCo Design Study
€10M | 2014 - 2017

DiSSCo Deploy
2024 - 2025

DiSSCo Construct
2021 - 2024

DiSSCo Prepare
€5M | 2019 - 2022

ENVRI
community
€19M | 2019-2022


ICEDIG
€3M | 2018-2020

Catalogue of Life
€0.8M | 2017 - 2020

SYNTHESYS
Synthesis of systematic resources
SYNTHESYS+
€11M | 2019 - 2021

mobilise
MOBILISE
€0.6M | 2018 - 2022



DISCO
Distributed System of Scientific Collections



 @catalogueoflife

 www.catalogueoflife.org



Catalogue of Life



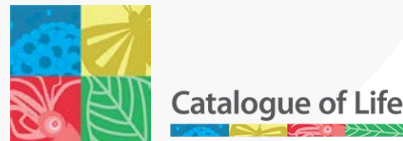
 - TAXONOMIC BACKBONE
Distributed System of Scientific Collections

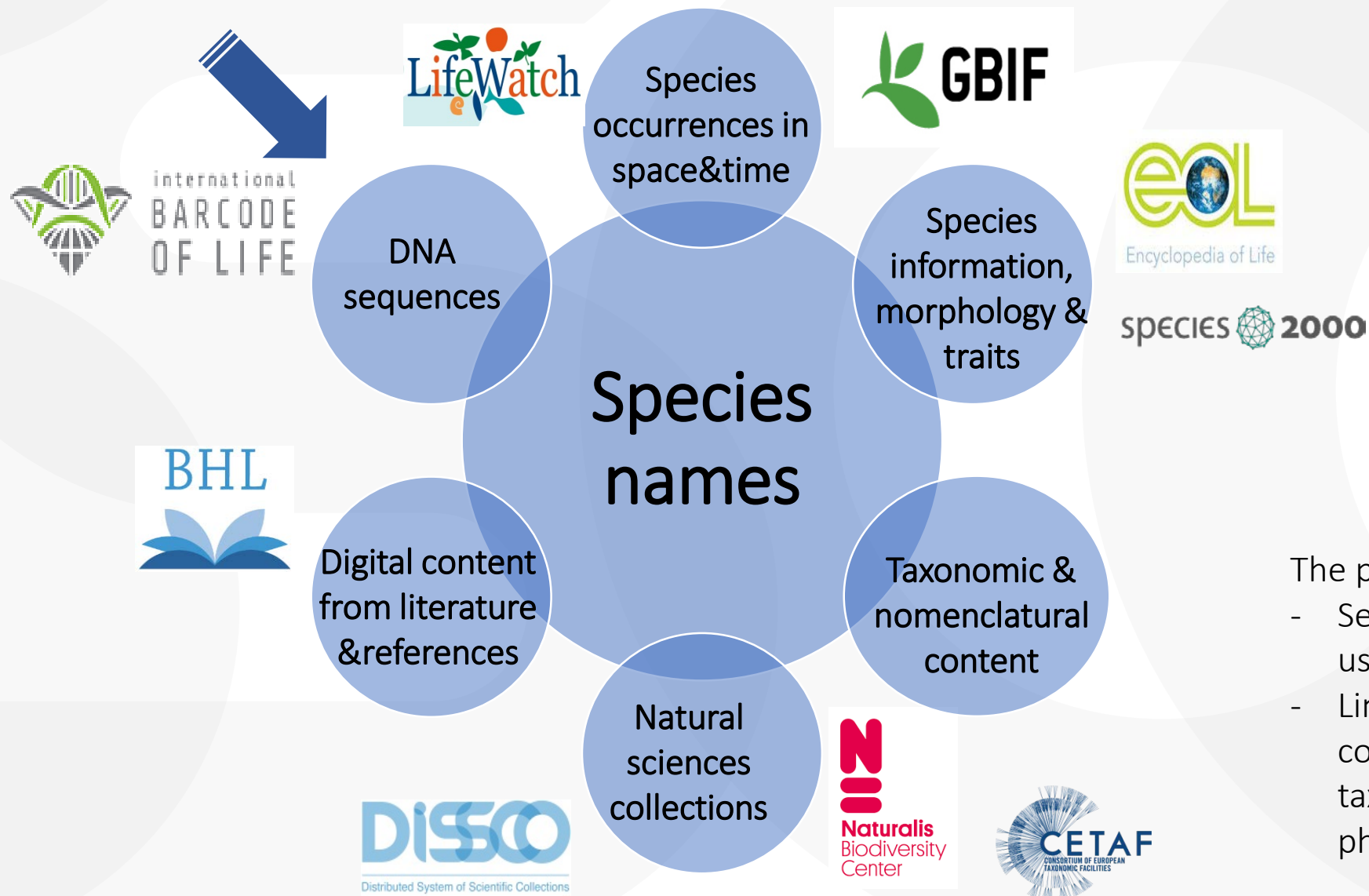
Catalogue of Life+

CoL+ will provide the taxonomic backbone for unified platform

CoL+ aims to:

- create an **open, shared, and sustainable consensus taxonomy and nomenclature foundation** to serve the proper linking of data in the global biodiversity information initiatives
- establish a clearinghouse for nomenclature and taxonomy to **reconcile sources**
- Build a comprehensive and authoritative **global index of species** (integrated species checklist and taxonomic hierarchy)





- The project works on
- Serving institutional users (1st phase)
 - Linking with taxonomic community to feed in taxonomic data (2nd phase)



@ENVRIcomm



www.envri.eu



DISCO - CLUSTER SERVICES
Distributed System of Scientific Collections

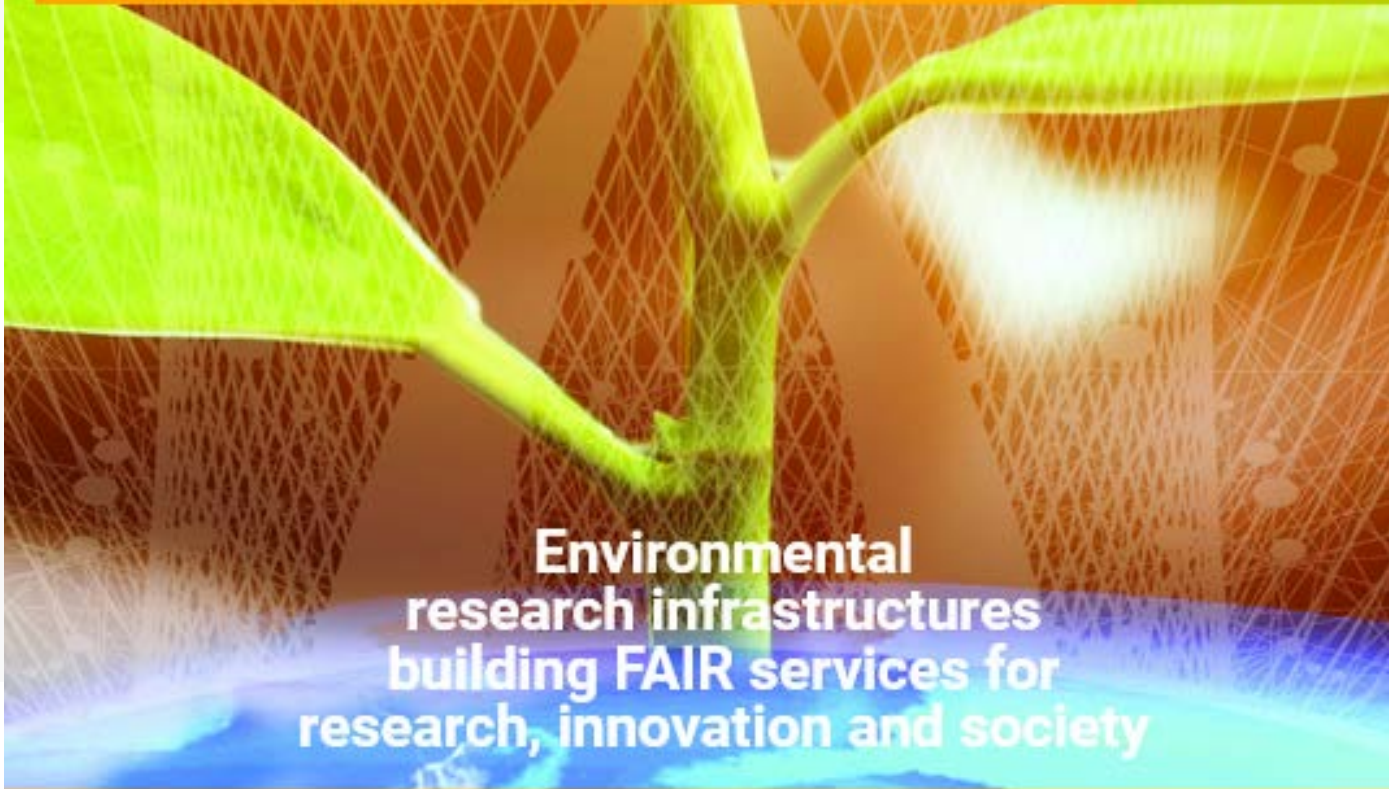
Environmental Research Infrastructures

ENVRI-FAIR is the connection of the Cluster of ENVRI to the EOSC

ENVRI+ objectives comprise:

- community **policies and standards** on all steps of the data life cycle
- data services compliant to the **FAIR principles**
- prototypes for testing **pre-production services** at each RI
- **thematic data services** and tools provided by the ENVRI





**Environmental
research infrastructures
building FAIR services for
research, innovation and society**

14

ENVRI-FAIR brings together 14 environmental research infrastructures

37

37 partners from 13 EU countries work together to make the project a success

48

The project has started in January 2019 and will run for 48 months until December 2022

12

Our work is organized in 12 work packages and will be documented in 80 deliverables

Reliable predictions of environmental change must be based on trustworthy, well-documented observations which capture the **entire complexity of the Earth system and the manifold interactions between the atmosphere, the land and the ocean**, encompassing impacts both from and onto life in all its forms.

ENVRI-FAIR will set technical preconditions for a virtual, federated machine-to-machine interface (ENVRI-hub) for accessing to **environmental data and services**



@mobilise_action



www.mobilise-action.eu

mobilise

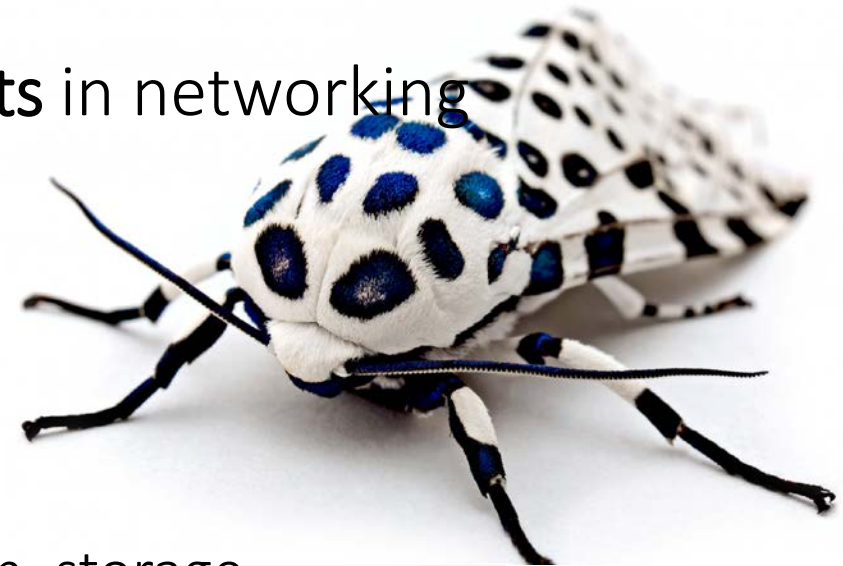
DISCO - COLLABORATIVE WORK
Distributed System of Scientific Collections

MOBILISE – CA 17106

MOBILISE for data, experts and policies in Scientific Collections

33 countries, 500 K€, 4 years and over 100 participants in networking and exchange

- Identifying **standards, processes and workflows, policies and technological gaps**
- Evaluating, **harmonising** and promoting those for acquisition, storage and publication of bio- and geodiversity data
- Creating an inclusive, **bottom-up and responsive network** of experts on bioinformatics



- community uptake
- comprehensive knowledge base
- knowledge and technology transfer



Training Schools

1st held on Digitization and Data Management Challenges of Small Collections
2nd will be held in February 2020 in Warsaw, PL.



Short Term Scientific Missions

For **early career scientists, collection and data managers, collection curators and stewards from smaller institutions** and institutions in countries that do not have extensive facilities or well-developed policies for data mobilisation, digitisation, data practices and curation

1ST MC1

Brussels, BE, 2 October 2018

2ND MC2, WGs and Training School Sofia, BU, 12-15 March 2019

3RD MC3, WGs and Training School Warsaw, PL, 10-15 February 2020



MOBILISE COST Action TRAINING SCHOOL 2019 14-15 March 2019
Sofia, Bulgaria



"Digitisation and data management challenges in small collections"

MOBILISE COST Action:
"Mobilising Data, Policies and Experts in Scientific Collections", in cooperation with GBIF <https://www.gbif.org> and iDigBio <https://www.idigbio.org> and under the umbrella of CETAF <https://cetaf.org> and DISCO <https://disco.wu>, organises its first Training School (TS) on the digitisation and management of biological and geological collections and the development of Digitisation Strategies & Plans.

Training School Overview
The TS will target a basic but important part of the digitisation procedure of biological and geological collections and will include the following topics:

- origins of data (taxonomic data, specimens and materials, literature data, field work notes, occurrence data, remote sensing data etc)
- data quality, ensuring the maximum quality when digitising taxonomic, geographical, collection and descriptive data
- data cleaning in order to improve the quality of data and make them "fit for use"

Most of the TS is dedicated to exercises on data quality and cleaning while group work will ensure and motivate interaction between trainees and trainers and foster lively discussions. Moreover, trainees will have the possibility to bring sample(s) of their institutional biodiversity and geodiversity digital database and exercise on them. Formal language of the TS is English.

Who could attend?
The Training School "Digitisation and data management challenges in small collections" is addressed to everyone who is engaged in biological and geological collections and their data, such as Curators and Collections' managers, Directors/Senior managers, Collections' Digitization managers/officers, Scientists on bio- or geo-informatics, Early Career Investigators (individuals who are within a time span of up to 8 years from the date they obtained their PhD/doctorate in full-time equivalent), Students (Post graduates, MSc, PhD), Technicians of collections. They must be engaged in an official research programme, or employed by, or affiliated to, an institution, organisation or legal entity, which has within its remit a clear association with performing research.

Location
The Training School will be held in Sofia, Bulgaria, from 14th to 15th of March 2019. It will be hosted by the Institute of Biodiversity and Ecosystem Research and the National Museum of Natural History - Sofia, both units of the Bulgarian Academy of Sciences.

Trainees Grants and Certificates
No fees apply to the Training School. Trainees will receive a fixed COST Grant amounting to 600 Euros. Certificates with ECVET points will be provided to trainees.

Registration
Applicants need to register here: <https://form.jotform.com/922281960390> and upload a CV of one page and a motivation statement of 200 words. The Training School is limited to 30 trainees. COST policies on gender balance, involvement of Early Career Investigators, and geographical inclusiveness will be followed. **Deadline of applications is the 10th of February 2019.**




MOBILISE COST Action 17106 SECOND TRAINING SCHOOL 13-14 February 2020
Warsaw, Poland



"Digitisation and data management challenges in small collections"

MOBILISE COST Action 17106
"Mobilising Data, Policies and Experts in Scientific Collections", in cooperation with GBIF <https://www.gbif.org>, CETAF-DEST <http://bascomtraining.eu> and DISCO <https://disco.wu>, organises its second Training School (TS) on the digitisation and management of biological and geological collections and the development of Digitisation Strategies & Plans.

Training School Overview
The TS will target the following basic but important steps of the digitisation process of biological and geological collections:

- The **data quality** issue, for ensuring the maximum quality when digitising taxonomic, geographical, collection and descriptive data, such as, specimens and materials, literature data, field work notes, occurrence data, remote sensing data, etc.
- The **data cleaning** in order to further improve the quality of data and make them "fit for use": by detecting and determining error types, search and verify error instances, correct the errors, document error instances and error types and modify data entry procedures to reduce future errors.
- The **data visualisation** focusing on applying visualisation techniques in clear data.


The TS will be a hands-on experience, since most of the TS is dedicated to exercises on data quality, cleaning and visualisation while group work will ensure and motivate interaction between trainees and trainers and foster lively discussions. Moreover, trainees will have the possibility to bring sample(s) of their institutional biodiversity and geodiversity digital database and exercise on them. **Formal language of the TS is English.**

Who can apply
The Training School is addressed to everyone engaged in biological and geological collections and their data such as Curators and Collections' managers, Directors/Senior managers, Collections' Digitization managers/officers, Scientists on bio- or geo-informatics, Early Career Investigators (individuals who are within a time span of up to 8 years from the date they obtained their PhD/doctorate in full-time equivalent), Students (Post graduates, MSc, PhD), Technicians of collections. They must be engaged in an official research programme, or employed by, or affiliated to, an institution, organisation or legal entity, which has within its remit a clear association with performing research.

Location
The two-days TS will be held in Warsaw, Poland on 13th and 14th of February 2020, hosted by the University of Warsaw.

Trainees Grants and Certificate
No fees apply to the Training School. Trainees will receive a fixed COST Grant amounting to 540 euros. Certificates with ECVET points will be provided to trainees.

Registration
Applicants need to register at the link <https://form.jotform.com/92101915982056> and upload a CV of one page and a motivation statement of 200 words. The Training School is limited to 30 trainees. COST policies on gender balance, involvement of Early Career Investigators, and geographical inclusiveness will be followed. **Deadline of applications is the 1st of December 2019.**



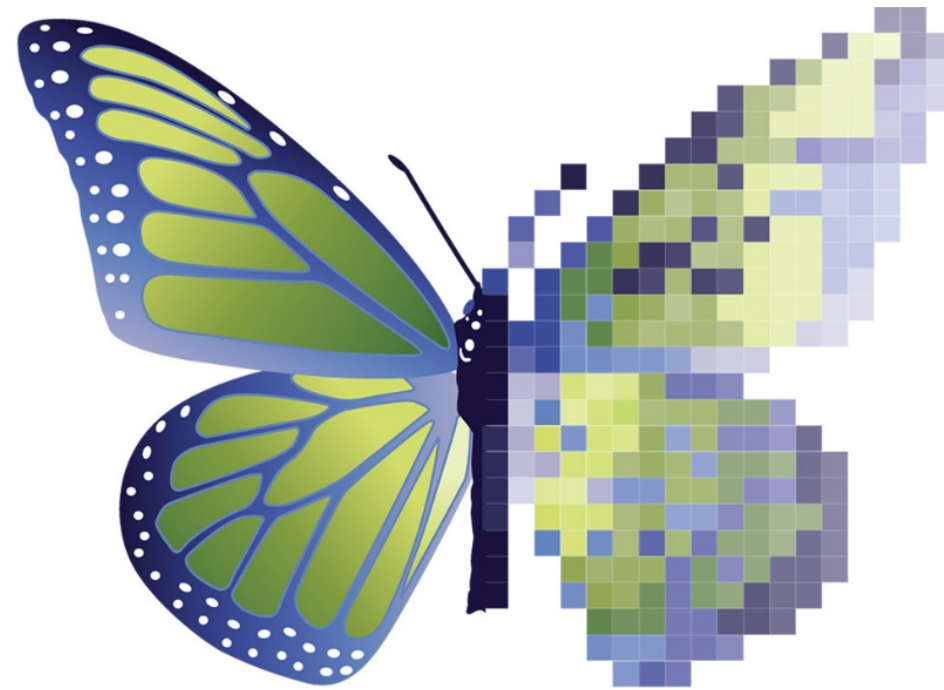

Centrum Nauk
Biologiczno-Chemicznych
Uniwersytetu Warszawskiego



@ICEDIG_eu



www.icedig.eu



ICEDIG.EU

DISCO
Distributed System of Scientific Collections

- DESIGN STUDY

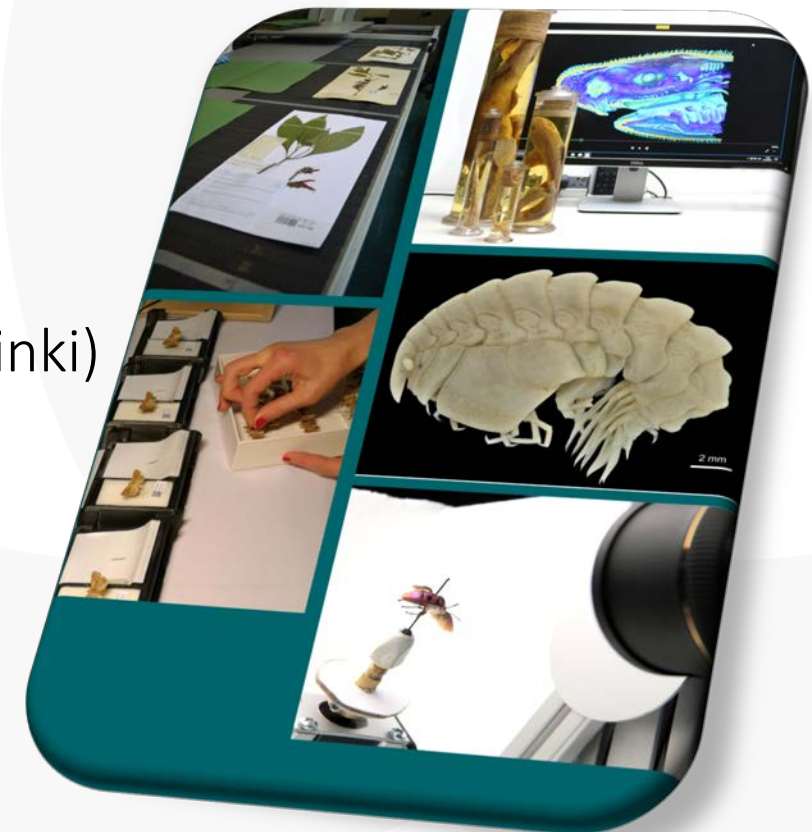
ICEDIG – Innovation and Consolidation for Large-Scale Digitisation of Natural Heritage



ICEDIG is the design study for DiSSCo

The project aims at refining the **technical, financial, policy and governance aspects of DiSSCo RI**

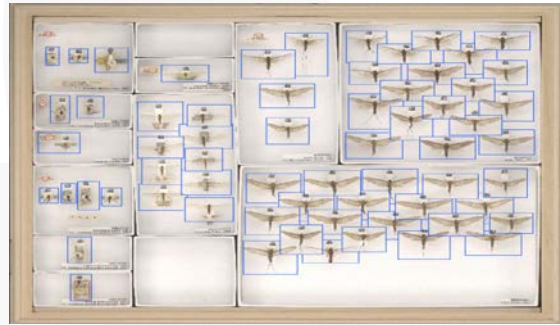
- **Partners:** 12 partners from 7 European countries (NHM Helsinki)
- **Duration:** 27 months (Jan2018-March2020)
- **Budget:** 3Mio €



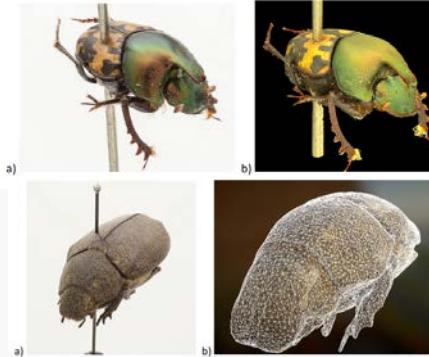
H2020-INFRADEV-2016-2017 – Grant Agreement No. 777483
Funded by the Horizon 2020 Framework Programme of the
European Union

Recent Outputs

State of the art in Digitization efforts



Pinned Insects



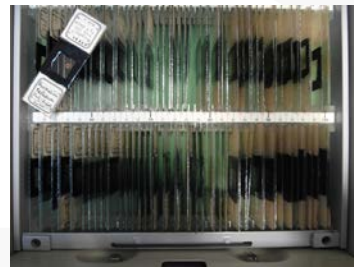
3D Capture



Liquid Samples



Skins and other vertebrate material



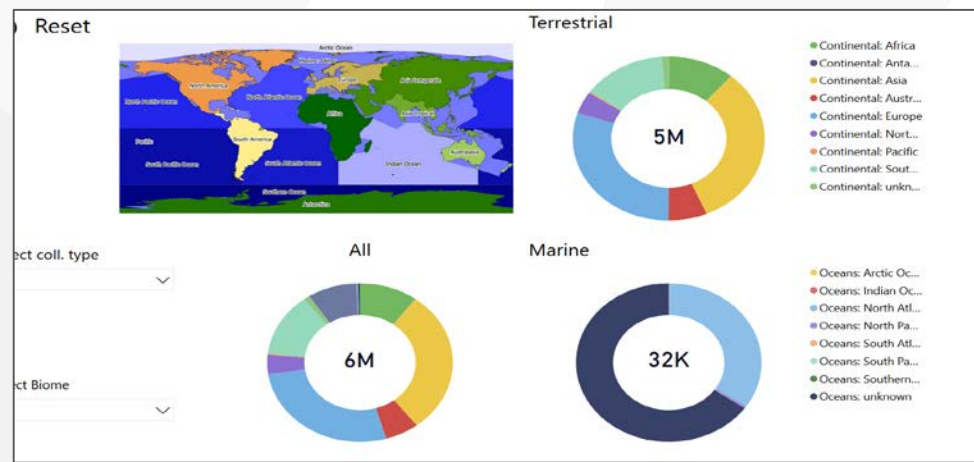
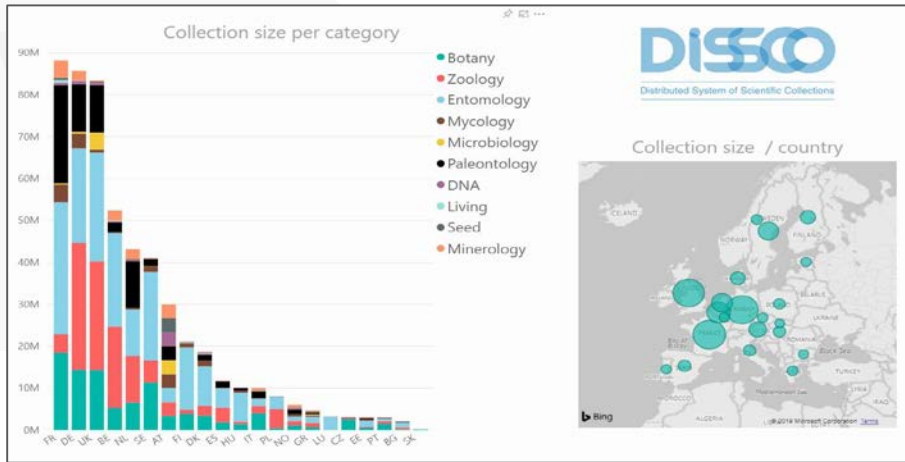
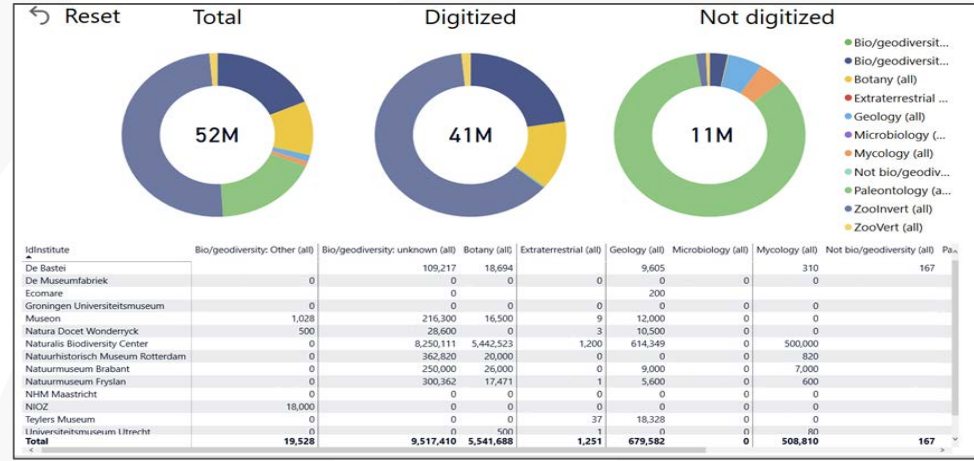
Microscopic Slides

Digitization Techniques, methods

Crowdsourcing, data quality and data standards...

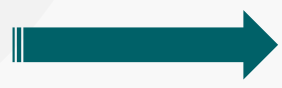
Training needs...

...as well as digitisation infrastructure design, open access, policies and standards



The CDD will provide information condensed in one view to enable taking **decisions both fast and informed**, based on accurate, reliable and persistent data on biodiversity-related issues at a global scale.

Digitization Dashboard



Instrumental mechanism for decision-taking

 @synthesysEU

 <https://www.synthesys.info>



 - UNIFY and INTEGRATE
Distributed System of Scientific Collections

SYNTHESYS+ of Systematic Resources

SYNTHESYS+ works to unify and integrate standards and pathways

The project aims to produce an accessible, integrated European resource based on a shared **management, preservation, and access**

- **Partners:** 32 partners from 15 European countries (NHM London)
- **Duration:** 48 months (2018-2020)
- **Budget:** 10 Mio €



Transnational (physical) Access to collections

- Provide funding for scientists to undertake **short visits to collections**
- Open collections, **staff expertise and analytical facilities** so that they become available for users, at any one of the 21 partner institutions

First Call closed in April '19

- **586** projects submitted, **7402** user days requested
- **74** non-EU applications
- Visits taking place throughout May



Virtual (digital) access to collections

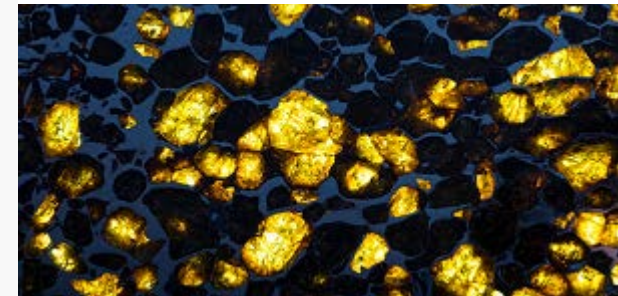
- Remove reliance on physical access by piloting digitisation workflows for DiSSCo
- Facilitate prioritising of DoD (Digitisation on Demand) and provide funding for digitisation of one or more collections

Call 1 of Virtual Access will run in Spring 2020

Right Now:

Soft 'ideas' Call to open discussion between researchers and institutions to shape proposals

More info at <https://www.synthesys.info/virtual-access.html>



Tools and methods for optimization of access, digitization on demand and specimen data refinery

ELViS

European loans and visit system

ELViS support loans, visits, applications for Access, and track outputs. Smoother, faster, and better access to natural history collections across Europe.

[Register now](#)



ELViS european loans and visit system

Institutions Requesters Collections Facilities Calls Requests

Jeremy Miller Requester

Search collections Advanced search 5 Selected

← Chrysops quadratus Meigen, 1820

Request

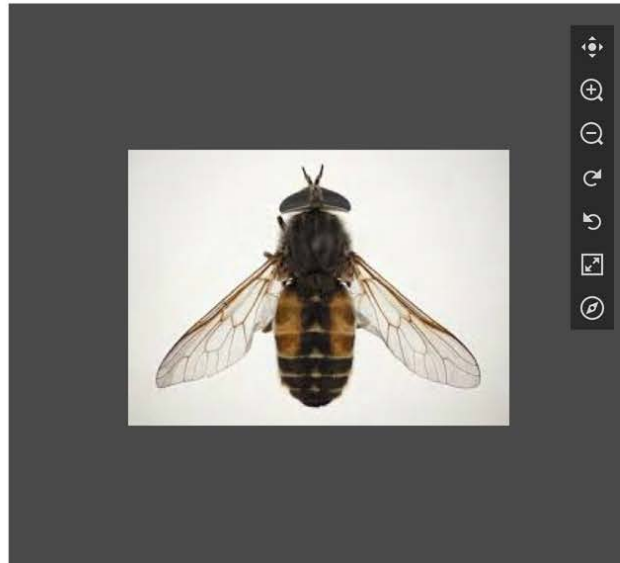
Details

Synonym of: [Chrysops viduatus \(Fabricius, 1794\)](#)

Classification: [Animalia](#) / [Arthropoda](#) / [Insecta](#) / [Diptera](#) / [Tabanidae](#) / [Chrysops](#)

Institute

Name: [Naturalis Biodiversity Center \(NL\) - Diptera](#)



Performance

Readiness

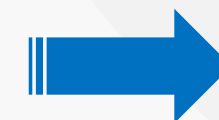
Specialisation

Priorities



In SYNTHESYS-previous iterations researchers and amateurs from over 40 countries spent approximately **9,200 visitor days** in the research departments and libraries.

Thanks to national projects (as Global Plants Initiative, REFLORA, the Virtual Herbaria portal and others), digitization of botanic collections is a well recognized reference. NHMW provides virtual access to the digital Herbarium which is used by more than **37** institutions from Afghanistan to El Salvador with up to **1,300** users per day



In SYNTHESYS+
> 25 PMs
> 420 K€

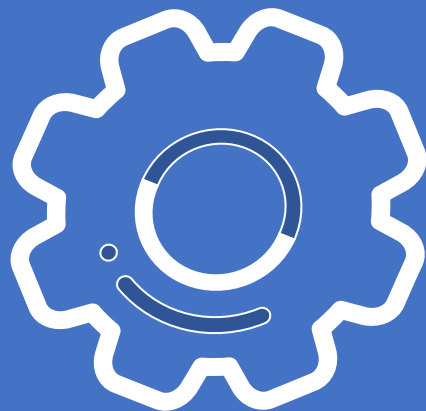
- Management (NA1)
- Training and support – CETAF (NA2)
- Molecular standards and processes – GGBN (NA3)
- Digital standards and processes – TDWG (NA4)
- Internationalisation – GBIF (NA5)



**Common
digital
agenda**

 @DiSSCoEU

 <https://www.dissco.eu>



DiSSCO

P R E P A R E

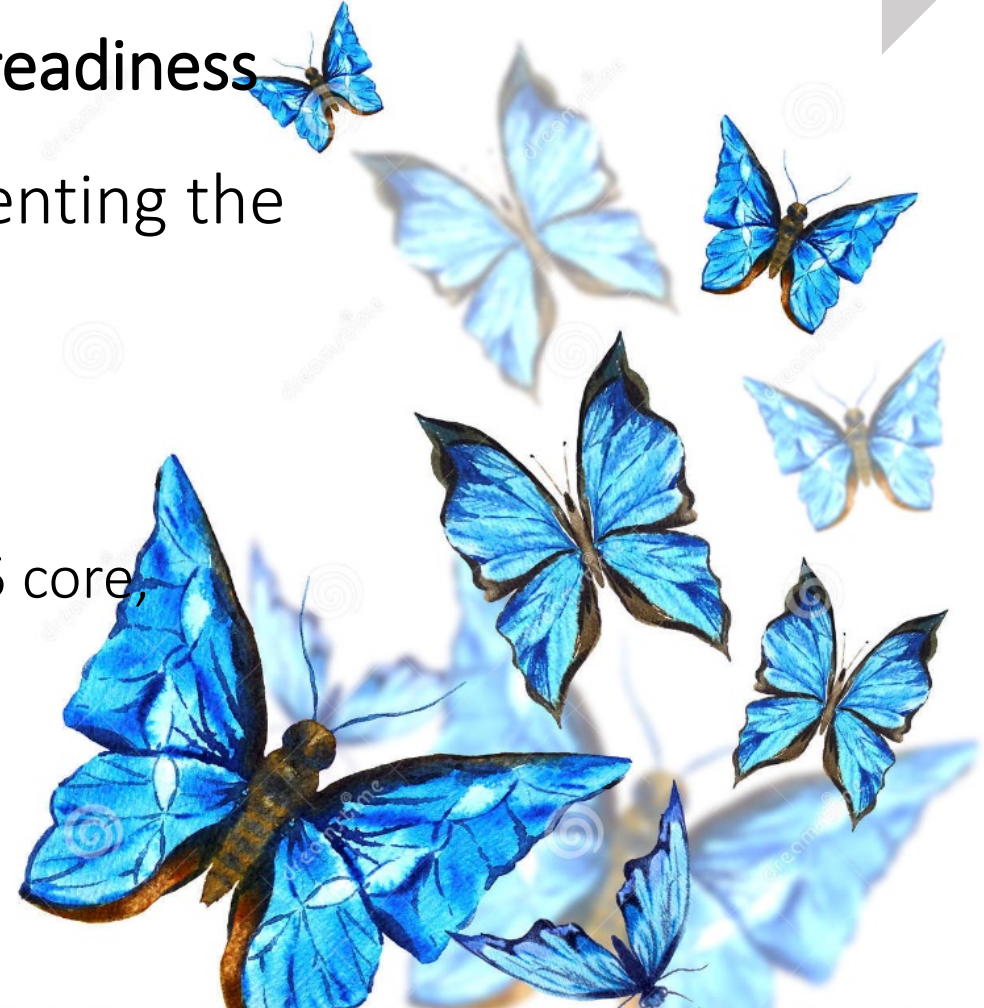
 - MASTER PLAN
Distributed System of Scientific Collections

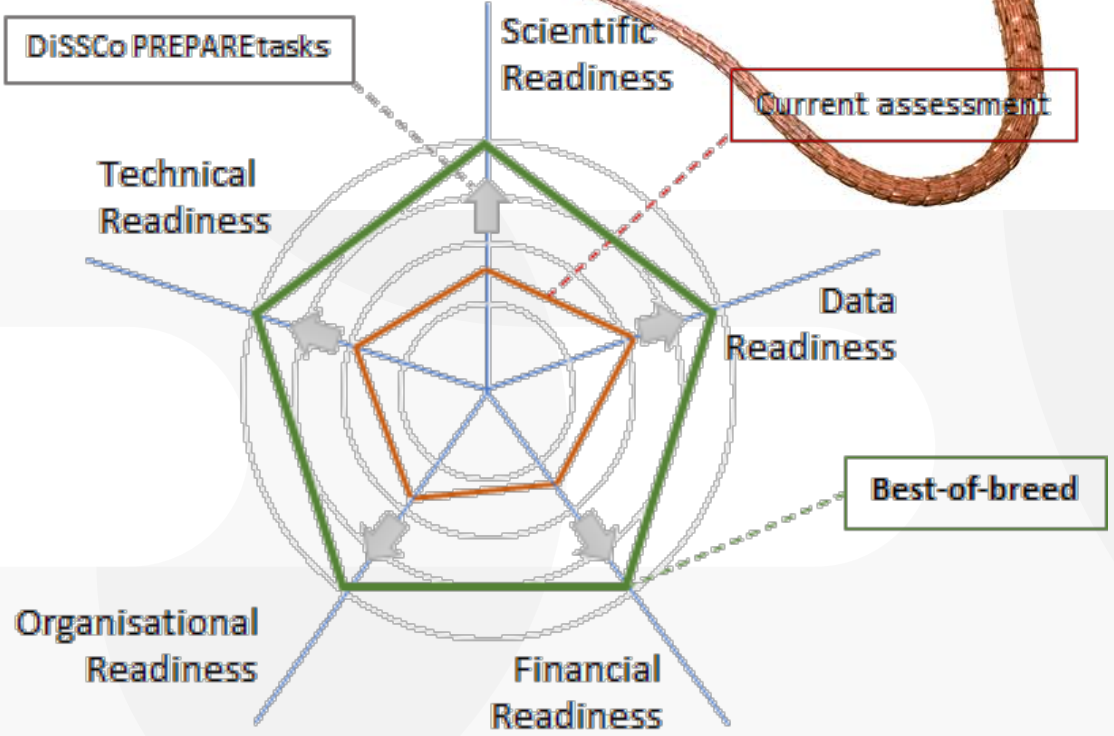
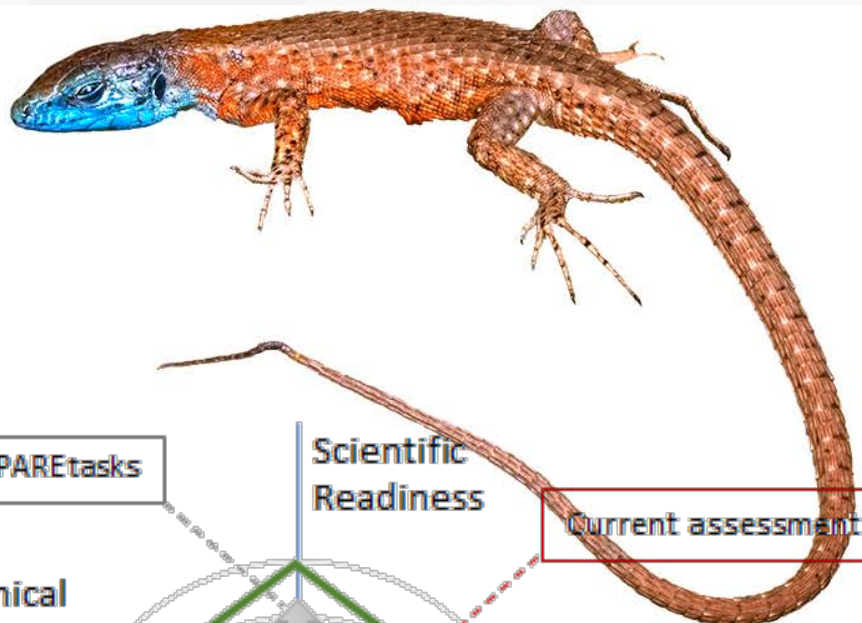
DiSSCo PPP – Preparatory Phase

DiSSCo Prepare will provide the Master Plan

It will facilitate achieving the necessary **level of readiness (IRL)** and will set up the **Master Plan** for implementing the construction and further operation

- **Partners:** 31 partners from 21 European countries (15 core, 4 technical and 12 National representatives)
- **Duration:** 3 years (2020-2022)
- **Budget:** 4,9 Mio € (incl. EU funds of 3,9M€)





| Dimension | Acronym | Definition |
|--------------------------|---------|---|
| Scientific Readiness | SR | Capacity of the RI to respond to/adjust against current and anticipated user needs |
| Data Readiness | DR | Capacity of the RI data producers and stewards to serve FAIR data |
| Technological Readiness | TR | Capacity of the RI to meet the functional requirements of its user audience through comprehensive and sustainable technological solutions |
| Financial Readiness | FR | Capacity of the RI to put in place and implement a comprehensive business model |
| Organisational Readiness | OR | Capacity of the RI to set and implement fit-for-purpose governance and management policies as well as strategic and operational plans |

Users needs & Socio-economic impact

HRs, Training and users support

Capacity enhancement

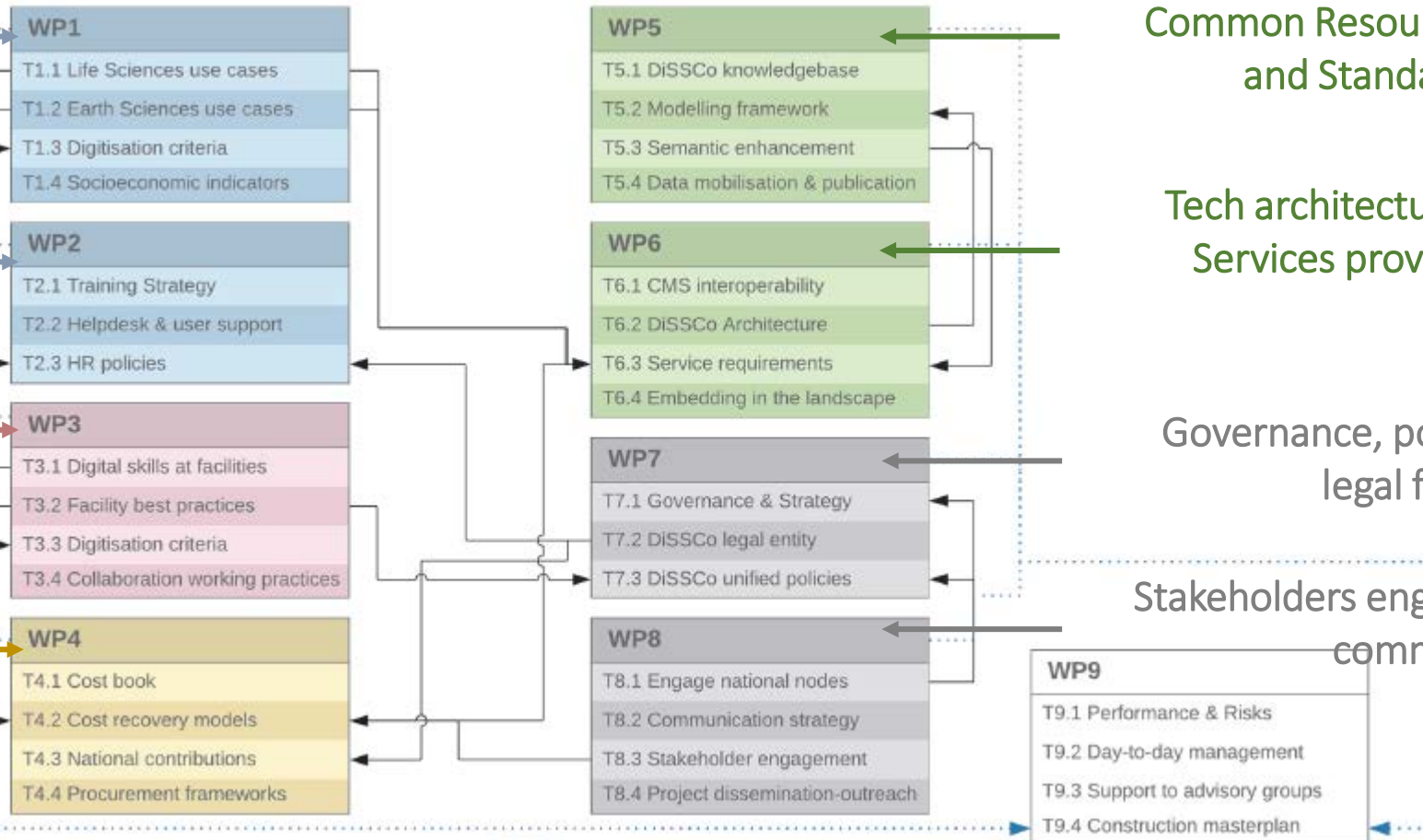
Business Framework

Common Resources and Standards

Tech architecture & Services provision

Governance, policy & legal frames

Stakeholders engage & community





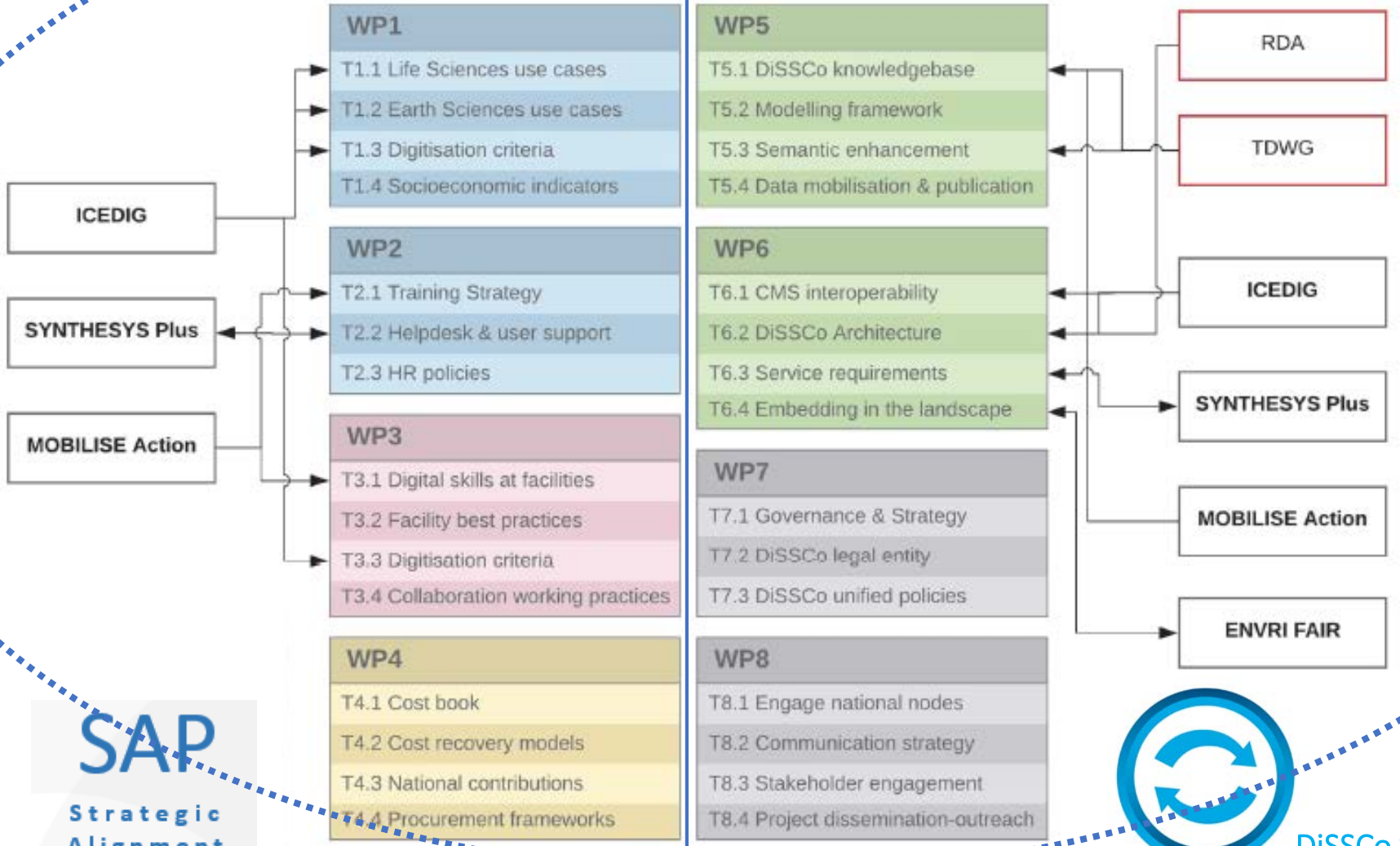
DiSSCo CSO



TAB



SAB



SAP
Strategic Alignment of Projects

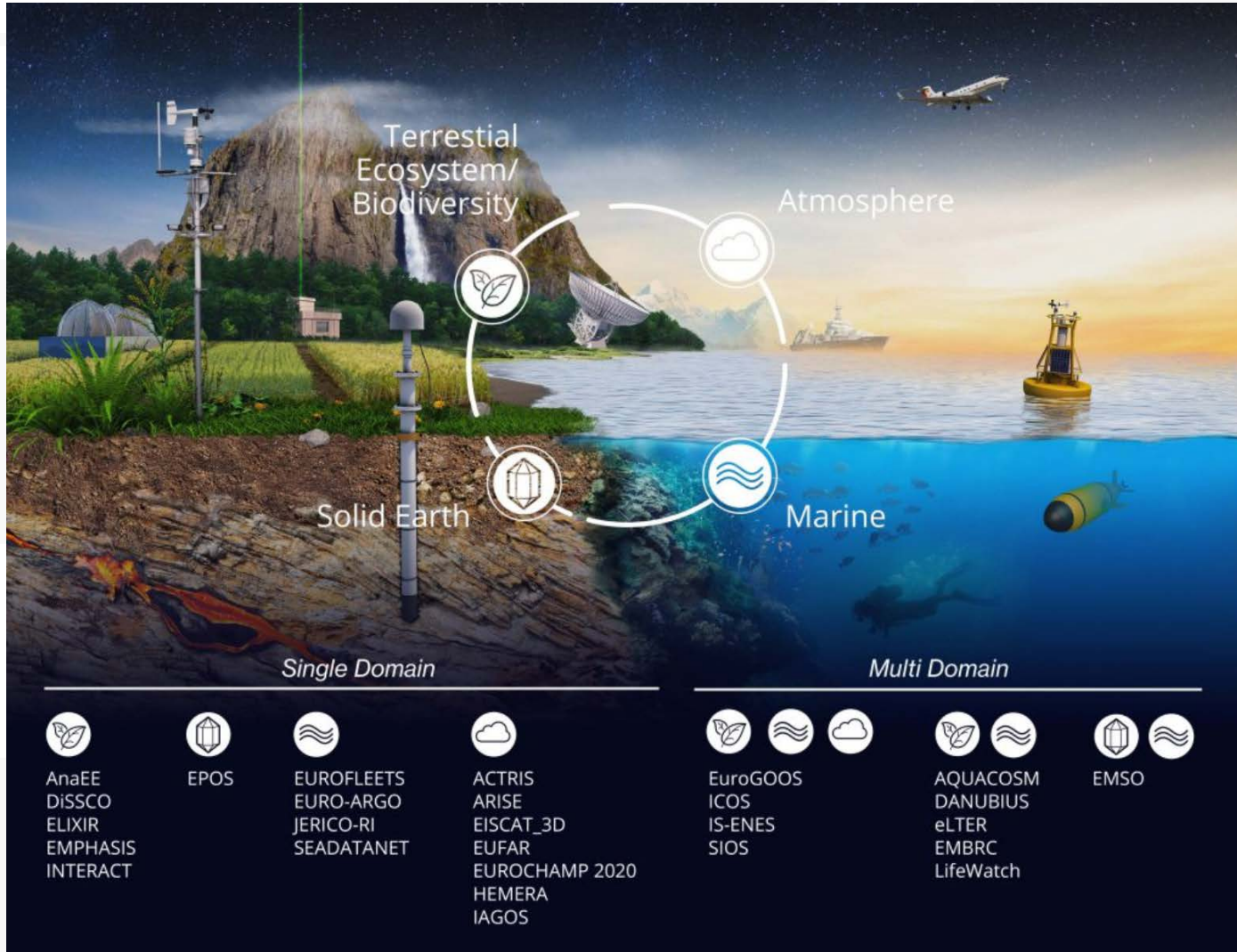


DiSSCo Synchronization Groups

Effective national participation

- Continue building a national consortium
- Call for Applications for membership of the DiSSCo GA
 - Launched in coming days
 - Submissions by 14 February 2020
- Kick-Off Meeting
 - Helsinki, FI, 27 and 28 February 2020
- General Assembly
 - Helsinki, FI, 27 and 28 February 2020

What Next?



In the DOMAIN

Collaboration via:

- Fora (BEERi)
- Projects (ENVRI-FAIR)

Linkage for leverage and
multi-purpose services

What
Next?

DISCO
Distributed System of Scientific Collections

CETAF
CONSORTIUM OF EUROPEAN
TAXONOMIC FACILITIES



OUT OF THE BOX

In Cultural Heritage

- Other RIs (E-RIHS, DARIAH)
- Repositories (Europeana)
- Modelling tools (Time Machine)

Creating a **shared think tank** for
Joint developments

What Next?

DISCO
Distributed System of Scientific Collections



GLOBALLY

Collaboration via:

- Aggregators (GBIF)
- Sister initiatives (iDigBio)
- Data Alliance (RDA)

Harmonized **standardization**
and **complementarities**

NETWORKING

Via CETAF and COST Actions (as MOBILISE and BioMONITORING):

- In the Working Groups
- At relevant International (IPBES, CBD, GTI, CSP)
- To developing a joint research agenda
- To provide resources (CETAF stable identifiers), best practices (e-publishing) and policies and papers (Bratislava Declaration)
- To build required services for the community (tools, training)

Strengthening the community to be an **independent and sound** scientific voice in a diversity of areas



DiSSCo

Distributed System of Scientific Collections



Thanks for your
attention!

Ana Casino
CETAF Executive Director
DiSSCo Deputy Coordinator

ana.casino@cetaf.org



DiSSCo
Distributed System of Scientific Collections