

## 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](mailto:info@dissco.eu).

### Title

DPP Deliverable 8.2 Thematic Specialisation Plan, Prototype tool

### Author(s)

Patrick Semal, Carole Paleco, Serge Scory (RBINS), Franck Theeten (RMCA/CETAF), Katharine Worley (MNHN), Lorenzo Cecchi, Gianna Innocenti (Unifi), François Dusoulier (MNHN), Ana Casino (CETAF), Isabel Rey, Celia Santos (CSIC), Aino Juslen, Kari Lahti, Heli Fitzgerald (LUOMUS), Eva Alonso, Niels Raes (Naturalis), Maria Joao Fonseca (MHNC).

### Identifier of the author(s)

Patrick Semal (0000-0002-4048-7728), Carole Paleco (0000-0002-2981-5436), Franck Theeten (0000-0002-0626-538X), Serge Scory (0000-0003-2692-8651), Katharine Worley (0000-0003-2377-6840), François Dusoulier (0000-0001-9062-5239), Ana Casino (0000-0002-9869-6573), Aino Juslen (0000-0001-9434-5250), Kari Lahti (0000-0003-0440-2069), Heli Fitzgerald (0000-0002-6754-6409), Eva Alonso (0000-0001-5336-9723), Niels Raes (0000-0002-4329-4892), Isabel Rey (0000-0002-2122-5124), Celia Santos (0000-0001-9507-7657), Lorenzo Cecchi (0000-0001-9706-0947), Gianna Innocenti (0000-0002-4504-0765), Maria Joao Fonseca (0000-0002-5138-875X).

### Affiliation

Royal Belgian Institute of Natural Sciences, Brussels Belgium, Patrick Semal, Carole Paleco, Serge Scory

Royal Museum for Central Africa, Belgium: Frank Theeten

Università degli Studi di Firenze: Firenze, IT: Lorenzo Cecchi, Gianna Innocenti.

Museo Nacional de Ciencias Naturales, ES: Isabel Rey, Celia Santos

Consortium of European Taxonomic Facilities (CETAF): Ana Casino

Muséum national d'Histoire naturelle (MNHN): François Dusoulier and Katharine Worley

Finnish Museum of Natural History (LUOMUS): Heli Fitzgerald, Kari Lahti, Aino Juslen

Naturalis Biodiversity Center, NL: Eva Alonso, Niels Raes

### Contributors

Universidade do Porto, PT: Maria Joao Fonseca

**Publisher**

**Identifier of the publisher**

**Resource ID**

**Publication year**

2023

**Related identifiers**

**Is it the first time you submit this outcome?**

Yes

**Creation date**

30/01/2023

**Version**

1

**Citation**

Semal, P., Paleco, C., Theeten, F., Scory, S., Alonso, E., Casino, A., Dusoulier, F., Worley, K., Cecchi, L., Innocenti, G., Rey, I., Santos, C., Juslen, A., Lahti, K., Fitzgerald, H., Raes, N., Fonseca, MJ.. Thematic Specialisation Plan, Prototype Tool. DiSSCO Prepare WP8 - Deliverable 8.1

**Abstract**

The Distributed System for Scientific Collections (DiSSCo) Research Infrastructure (RI) will operate a number of e-services. This deliverable under Task 8.1 proposes to add a Tool to these services to be managed by DiSSCo RI, and to be continuously fed by DiSSCo Institutions and National Nodes (NN). Its main asset is to generate search outputs highlighting the specialisations and expertise existing among the community of DiSSCo RI users and content providers. The Specialisation tool can provide illustrations, schemes and lists of measurable data on the Natural Sciences Collections (NSC), on the community's assets for collection management and digitization.

**Content keywords**

data

**Project reference**

DiSSCo Prepare (GA-871043)

**WP number**

WP8

**Project output**

Deliverable

**Deliverable/milestone number**

Deliverable 8.2

**Dissemination level**

Public

**Rights**

**License**

CC0 1.0 Universal (CC0 1.0)

**Resource type**

Text

**Format**

pdf

**Funding Programme**

H2020-INFRADEV-2019-2

**Contact email**

[carole.paleco@naturalsciences.be](mailto:carole.paleco@naturalsciences.be)



H2020-INFRADEV-2019-2  
Grant Agreement No 871043

# Thematic Specialisation Plan, prototype tool

## DiSSCo Prepare WP8 – Deliverable 8.2

Authors: Patrick Semal, Carole Paleco, Serge Scoíy (RBINS), Katharine Worley, François Dusoulier (MNHN), Ana Casino (CETAF), Aino Juslen, Kari Lahti, Heli Fitzgerald (LUOMUS), Eva Alonso, Niels Raes (Naturalis), Isabel Rey, Celia Santos (CSIC), Lorenzo Cecchi, Gianna Innocenti (Unifi), Maria Joao Fonseca (MHNC), Franck l'heeten (RMCA/CETAF).



## Abstract

The Distributed System for Scientific Collections (DiSSCo) Research Infrastructure (RI) will operate a number of e-services. This deliverable under Task 8.1 proposes to add a Tool to these services to be managed by DiSSCo RI, and to be continuously fed by DiSSCo Institutions and National Nodes (NN). Its main asset is to generate search outputs highlighting the specialisations and expertise existing among the community of DiSSCo RI users and content providers. The *Specialisation tool* can provide illustrations, schemes and lists of measurable data on the Natural Sciences Collections (NSC), on the community's assets for collection management and digitization.

## Contribution to DiSSCo RI

National Smart Specialisation and institution-level strategies inform prioritisation objectives in each DiSSCo member country. Gathering the state-of-the-art nationally will **provide the basis for the construction of an overall strategic map that is necessary for DiSSCo activities distribution and operation**. Equally important will be to **establish a follow up mechanism** to ensure alignment and harmonisation with national RI roadmap processes and relevant foreseen developments (e.g. national contributions/nodes for the European Open Science Cloud (EOSC), participation in cluster services development, etc.). The Specialisation tool is aimed at being a centralised platform developed under the DiSSCo Prepare (DPP) project to ensure that in addition to the main digitization and management techniques and processes, the documentation related to the national strategies, centres of excellence and services providers will be accessible through a single entry point, and will be part of the searchable information available for DiSSCo users and stakeholders looking for key technical and scientific assets.

## Keywords

Specialisation, collections, training, facility, facilities, expertise, galleries, scientific instrument, specimen, digitization, staff, curation, MIDS, data.

## Table of Contents

1. Introduction and Objectives	4
2. Development of the tool	6
2.1 Methodology	6
2.2 Design	6
2.2.1. Design of the user interface of the Specialisation tool using Google forms	8
2.2.2. Modifications of the forms using Google scripts	14
2.2.3. Access to the Google forms and to the Google sheets under the NNs spaces	16
2.2.4. Aggregation of the data using Google scripts	18
2.2.5. Creation of graphs and maps for the analysis of the data	20
2.2.6. Data analysis and search	27
2.2.8. Update of the information	31
2.3 Knowledge base Integration – ( links with existing dashboards )	33
2.4 Software code	33
3.1 Data collection methodology	34
3.2 First outputs from current exercise	35
4. User Manual	37
5. Recommendations	38
5.1 Future use of the tool	38
5.2 Continued data compilation	38
6. References	40
7. Appendix : User Manual	40

## 1. Introduction and Objectives

The Natural Science Collections (NSC) holding institutions gathered around CETAF (Consortium of European Taxonomic Facilities) constitute a comprehensive, stable and cohesive community that, among a diverse realm of initiatives, has promoted, supported and is currently building the Research Infrastructure named DiSSCo (Distributed System of Scientific Collections) on which to anchor the digital transformation that is taken place worldwide. In this perspective DiSSCo aims to facilitate virtual access to the incredible patrimony that those collections form and provide services that would allow effective exploitation of such precious resources.

The research activity around NSC is constantly evolving and the community is becoming progressively more mature both, technology and management-wise within the landscape of the European Research Area (ERA). The multiple facets and research activities linked to the collections are fostering the development of instruments, new skills and competencies enabling an easier access to the complex realm of specimen conservation, specimen management and related knowledge transfer as a precious heritage to understand biodiversity. This deliverable (D8.1) introduces a new tool that has been developed based on the existing portal Synthesys + Collections Digitisation Dashboard (CDD)<sup>1</sup> and the CETAF Passport and Collections Registry<sup>2</sup>.

Under DiSSCo Prepare WP8, Task 8.1 was aiming at providing a *Thematic Specialisation Plan* which would gather the NSC state-of-the-art and provide the basis for the construction of a strategic mapping required for the better understanding of the distribution of DiSSCo RI activities and its operation across European institutions. In order to develop such a *Thematic Specialisation Plan* the need for a repository able to gather which would then feed the data had to be created. The *Specialisation tool* proposes to collect key data from the NSC hosting institutions to meet the categories that had already been identified for the Collections Registry but also to more relevant categories thanks to the work done by the partners of the task. The *Specialisation tool* includes a follow up mechanism to ensure alignment and harmonisation with national RI roadmap processes and relevant foreseen developments (e.g. national contributions/nodes for EOSC, participation in cluster services development, etc.).

Thanks to the monthly exchanges under DPP WP8 with the representatives of the National Nodes, the development and objectives of the tool were presented to a large panel of future DiSSCo RI contributors and users throughout 2021-2022.

This deliverable presents the functioning and purposes of the tool, how the data are being gathered to create a large dataset of information about DiSSCo National Nodes, or institutions in case the National Node is not yet constituted at national level, that include collections, their

---

1

<https://app.powerbi.com/view?r=eyJrIjoiM2QyNDlmNjEtOTk3Ni00NGYwLThiMDctZTA5MjY0Mjc0MjUwIiwidCI6IjczYTI5YzAxLTRlNzgtNDM3Zi1hMGQ0LWM4NTUzZTE5NjBjMSIsImMiOiJh9>

<sup>2</sup> <https://collections.naturalsciences.be/cpb/cetaf-passport-and-collections-registry-manual>

digitization level and techniques, their conservation practices, the management, maintenance and research being done on the collections, as well as the training provided at institutional level and exhibitions deployed using specimens of those collections.

The objective of this tool is to collect the most complete information about DiSSCo partners describing four major assets, namely the collections, the research, the expertise and the training/education activities.

The collected information allows to have different levels of analysis. The data encoded by institutions individually can be displayed and used for internal purposes to illustrate the state of the art of each participating institution. Furthermore, the institutional data is aggregated allowing each National Node to manage the information and to produce statistics and figures at national level that can demonstrate several facets at country level such as the joint level of maturity of national NSCs, integration of assets into national roadmaps, and development of necessary supporting actions in a harmonised and structured manner. Similarly, the tool can allow comparison and analysis to indicate strengths and opportunities, to detect gaps and support prioritisation. The Specialisation tool thus becomes a powerful mechanism for policy makers in each country to take well-grounded measures in the field of NSCs.

Furthermore, the tool enables the aggregation of the information at infrastructure level allowing the analysis of the data at the overall DiSSCo community level. In this case, the tool can facilitate the decisions made around topics such as the identification of niches for action, the definition of centres of excellence and the coherent and comprehensive development of the infrastructure at large, to ensure its effective operation, to enlarge and improve its catalogue of services, and to serve to the largest groups of targeted users. Multiple searches applied to the data can result in sets of information on each National Node which can be reported as the specialisations emerging from the community. The tool has therefore been developed to enable the visualisation of specialisations existing among DiSSCo institutions and to integrate dashboards of assets at different scales and for multiple purposes.

The deliverable also provides, thanks to the first data delivered by the National Nodes members of DiSSCo Prepare, some scenarios/case studies displaying the potential of the tool as a search engine.

A first User Manual is provided in the corresponding Appendix to guide throughout the forms and categories aimed for data collection. The forms and categories have been developed specifically to correspond to the specificities of the natural science collections held by the institutions participating in the DiSSCo community. They have also taken into consideration the need for harmonisation of data to aggregate them in a meaningful way and to display results coherently. The categorization of collections is based on the work developed under Synthesys+ project (for the metadata schema supporting classification of specimens) and further enhanced and adapted to comply with specific requirements set up by the ELViS mechanism under DiSSCo RI. The Specialisation tool will be the feeding source of reliable information to this entry point. ELViS will provide a unified way to request visits, loans and virtual access (VA) to collections, the latter providing digitisation on demand as a new type of access. Thus, the Specialisation tool will contribute to ensure comprehensiveness, reliability and sustainability of ELViS.



Finally, a set of recommendations are included that outline the future direction for the development and use of the Specialisation tool as a strategic and powerful mean for DiSSCo RI.

## 2. Development of the tool

### 2.1 Methodology

The *Specialisation tool* is based on existing tools and standards developed under ICEDIG and Synthesys+ projects, as well as the Biodiversity Information Standards (TDWG) Collection Descriptions (CD) Interest Group, As well as the CETAF Collections Registry, and the aggregating initiative of GRSciColl undertaken by GBIF (the Global Biodiversity Informatics Facility), which respective details can be found in the following links:

- [ICEDIG deliverable D2.3 Design of a Collection Digitisation Dashboard Deliverable D2.3](#)
- [Synthesys + D2.2 JOINT DASHBOARD OF COLLECTIONS ASSESSMENT TOOLS](#)
- [CETAF Passport and associated collection registry](#)
- [Global Registry of Scientific Collections](#)
- [Latimer Core: A new data standard for collection descriptions](#)

The goal was to create a web interface collecting the information at institutional level for most of the data and to individually for expertise, as a baseline for further aggregating information at national and infrastructure levels..

The requirements and the criteria that were to be met for the development of the tool were the following:

- Collaborative effort: the input can be shared between several users from the same institution
- No need to create and maintain hundreds/thousands of user accounts
- Each National Node has its own user account
- Restricted access limiting the impact of robots (and hackers)
- Access to the information at national level by each National Node
- Export of the collected data in a structured file like XLS or CSV.
- Connection to the indexes by external systems using API
- Possibility to import data produced by existing institution/national databases
- Open Source and/or free technologies
- End user manuals to fill the information
- Technical documentation for maintenance as a DiSSCo e-service

The setup of the tool went through several iterations of development supported by the feedback received from the partners of Task 8.1 in DiSSCo Prepare (DPP).

### 2.2 Design

Technologies are constantly evolving which makes it even more difficult to choose a specific technology without the risk that this technology will not be supported in the near future. This is

the reason why we decided to use robust and existing technologies which are already used by DiSSCo institutions in the long term.

We preferred to combine several “standard” technologies rather than using a specific tool with the risk of disappearance in the next few years.

The CETAF Collections Registry is originally based on the Plone OpenSource Content Management System (CMS). Although Plone might not be the most used CMS compared to Drupal or Wordpress, it is recognized as the most stable and secured one. Since 2003, only 89 Common Vulnerabilities and Exposures were identified in Plone products against 2629 for Wordpress or 1074 for Drupal. Plone includes an integrated Object-Oriented Database allowing users to define with Graphical User Interface (GUI) new content types and fields. The Royal Belgian Institute of Natural Sciences (RBINS) has been using Plone CMS since 2005 and continues to develop new tools related to the collections and the online publication of data.

**Royal Belgian Institute of Natural Sciences**  
by padmin — last modified Aug 23, 2021 03:18 PM — History

Identifications

**Unique Acronym**  
"ISO code of Country"+"Acronym of the Institution" e.g. BE-RBINS  
BE-RBINS

**ROR ID**  
e.g. 02y22ws83 for RBINS (https://ror.org/02y22ws83)  
02y22ws83

**ISNI (ISO certified global standard number)**  
e.g. 0000000121719581 for RBINS (https://isni.org/isni/0000000121719581)  
0000000121719581

**Wikidata ID**  
e.g. Q16665660 for RBINS (https://www.wikidata.org/wiki/Q16665660)  
Q16665660

A specific structure was created in Plone to build the CETAF Collections Registry and specific content types were created using the Dexterity tool. The structure is composed of a hierarchy of folders with restricted access to public or private pages and files.

**Edit NH Institution**  
None

Default - Identifications Dashboard Annual Reports Categorization

**Unique Acronym**  
"ISO code of Country"+"Acronym of the Institution" e.g. BE-RBINS

**ROR ID**  
e.g. 02y22ws83 for RBINS (https://ror.org/02y22ws83)

**ISNI (ISO certified global standard number)**  
e.g. 0000000121719581 for RBINS (https://isni.org/isni/0000000121719581)

**Wikidata ID**  
e.g. Q16665660 for RBINS (https://www.wikidata.org/wiki/Q16665660)

**GRID ID**  
e.g. grid.20478.39 for RBINS (https://www.grid.ac/institutes/grid.20478.39)

**GRSciColl code**  
e.g. RBINS for Royal Belgian Institute of natural Sciences

**International Name**  
English Name of the institution

Moreover, a specific interface was developed to enable a manual input of the information. Taking into account that some DiSSCo users might not be very familiar with Plone CMS, and that the maximum number of connections for edition is limited to 16 users simultaneously, we decided to switch the input interface to a Google Form system.

Figure 1

### 2.2.1. Design of the user interface of the Specialisation tool using Google forms

Among the possible choices available, the developer team (RBINS/CETAF) has decided to create the User Interface of the Specialisation tool in Google Forms as these are already well-known and widely used by the DiSSCo community.

The first version of the Google Form interface was created on the basis of the Plone-based Collections Registry of CETAF. The Google Drive storage space was divided into specific sub-folders each related to a certain category and the specific forms that were created under each category.

The table below lists the 10 categories covered and the related Google Forms for data integration:

<b>Folders per category</b>	<b>Forms</b>
Collections	a) The list of the collections housed by the institution b) The detailed information about the collections (related to 12 main collections and 73 sub-collections)
Galleries	The description of the galleries that are displaying specimens of collections.
Exhibitions	The description of the exhibitions developed by the institution and showing the capacity to develop temporary exhibitions including collections.
Facilities	The facilities present and accessible in the institution.
Scientific Instruments	The scientific Instruments present and accessible in the institution.
Research fields	The research fields covered by the researchers from the institution. Linked to the institution's research strategy.
Trainings	The trainings proposed by the DiSSCo partners and offered to users.
Institution Expertises	The expertise specific to the institution. Ex Center of Excellence.
Individual Expertises	The expertise covered by the individual scientists of the institution.

This first version of the Specialisation tool was proposed and presented to the Task 8.1 partners for testing and feedback at the end of 2021.

Based on the discussions held and the comments made to the first, a second version of the Specialisation tool was developed to integrate major requests for improvement..

Those requests were:

- to simplify the selection of the Institution, limiting the choice by country;
- to manage the information at NN level(s);
- to enable the data encoding using XLS template and Excel as supporting applications.

The different Google Forms were duplicated for each country thus allowing the NN to manage the information at national level. There are therefore 10 specialisation Google Forms per country corresponding to the categories defined in the table above.

The Specialisation tool is thus composed of 230 (10 forms x 23 countries) Google Forms. The naming of the form is standardised using the ISO code of the country (ex. BE for Belgium) and the name of the category as specialisation form e.g. “BE Collections Full Information”.

When possible, we also created a XLS template corresponding to each Google Form allowing the institutions or National Nodes to fill the information using a spreadsheet, or sending their own to the collector (i.e RBINS/CETAF developer), instead of Google Forms. This option was created after the request of the Task 8.1 partners but is not considered as an optimal way to fill in the information as it requires an exhaustive and largely time-consuming post-treatment to incorporate the data and check the validity of the information.

It nevertheless has the advantage to include data exported from other databases managed at institutional or national levels, or already collected in the framework of previous exercises (e.g. Synthesys Digitization dashboard, One World Collection,...).

The image shows a Google Form titled "Identification" with a blue header. It contains four sections, each with a title and a description, followed by a text input field:

- Title (in English) \***  
Title of the Research field  
test NN Belgium
- Original Title**  
Title of the Research field in local language  
Your answer
- Summary (in English)**  
Short description to be displayed in results search table  
Your answer
- XLS File upload. \***  
1. Download the xls template available here: 2.Fill the requested information and 3. Upload the completed file using this File upload field.  
Add File

Figure 2

Upload of the information using an XLS template.

However, the XLS option is not possible for the 2 following forms:

- The **individual expertise**, as it contains personal data requiring the approval from each person (GDPR). The compilation by an Institution/NN Coordinator in an XLS file is therefore not possible.
- The **detailed information about the (sub)collection** as the form is composed of 214 possible fields and therefore not optimal for a manual editing of the XLS file.

For instance, the general MIDS values of a (sub)collection are integrated into the detailed information form. The specialised MIDS values, which represent a complete matrix of data, are normally produced by the Collection Management Systems (CMS) of the institutions. The XLS format is therefore more suitable than a manual input.

1	A	B	C	D	E	F	G
2	Stratigraphy unit	Object Quantity	Object quantity Confidence (%)	MIDS-0 (%)	MIDS-1 (%)	MIDS-2 (%)	MIDS-3 (%)
3	Phanerozoic/Any era						
4	Phanerozoic/Cenozoic/AnyPeriod						
5	Phanerozoic/Cenozoic/Quaternary/Any epoch						
6	Phanerozoic/Cenozoic/Quaternary/Holocene						
7	Phanerozoic/Cenozoic/Quaternary/Pleistocene						
8	Phanerozoic/Cenozoic/Neogene/Any epoch						
9	Phanerozoic/Cenozoic/Neogene/Pliocene						
10	Phanerozoic/Cenozoic/Neogene/Miocene						
11	Phanerozoic/Cenozoic/Paleogene/Any epoch						
12	Phanerozoic/Cenozoic/Paleogene/Oligocene						
13	Phanerozoic/Cenozoic/Paleogene/Eocene						
14	Phanerozoic/Cenozoic/Paleogene/Paleocene						
15	Phanerozoic/Mesozoic/AnyPeriod						
16	Phanerozoic/Mesozoic/Cretaceous						
17	Phanerozoic/Mesozoic/Jurassic						
18	Phanerozoic/Mesozoic/Triassic						
19	Phanerozoic/Paleozoic/AnyPeriod						
20	Phanerozoic/Paleozoic/Permian						
21	Phanerozoic/Paleozoic/Carboniferous						
22	Phanerozoic/Paleozoic/Devonian						
23	Phanerozoic/Paleozoic/Silurian						
24	Phanerozoic/Paleozoic/Ordovician						
25	Phanerozoic/Paleozoic/Cambrian						
26	Precambrian/Any era						
27	Precambrian/Proterozoic/AnyPeriod						
28	Precambrian/Proterozoic/Neoproterozoic						
29	Precambrian/Proterozoic/Mesoproterozoic						
30	Precambrian/Proterozoic/Paleoproterozoic						
31	Precambrian/Archean/AnyPeriod						
32	Precambrian/Archean/Neoarchean						
33	Precambrian/Archean/Mesoarchean						
34	Precambrian/Archean/Paleoarchean						
35	Precambrian/Archean/Eoarchean						
36	Precambrian/Hadean						
37							
38							

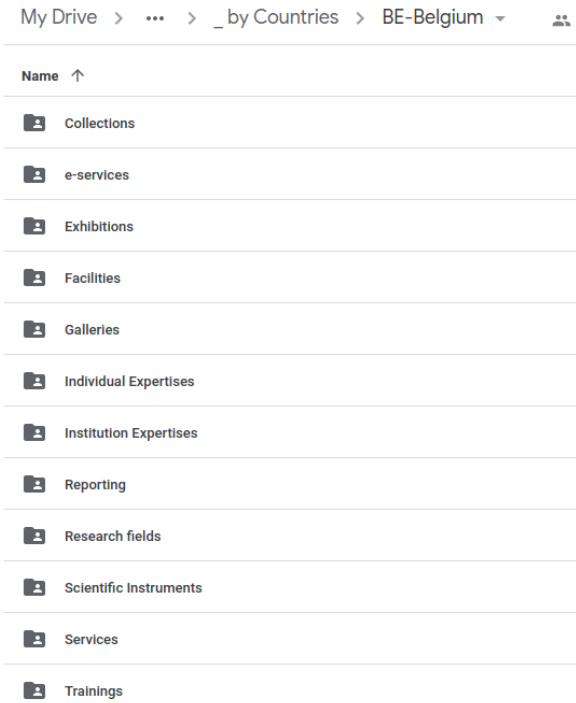
In fact, the MIDS values related to the Geography, the Stratigraphy and the Storage of the (sub)collections can only be entered through a proposed [XLS file](#), due to their specific dimensions that are not applicable generally.

Figure 3

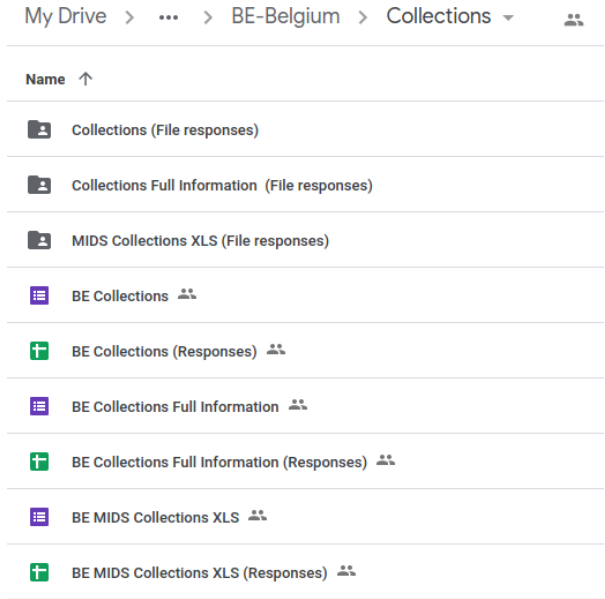
Once collected from the institutions, the Google Forms, the Google sheets with the collected data, and the uploaded XLS files are stored in a dedicated Google Drive with a folder hierarchy corresponding to the NN.

- AT-Austria
- BE-Belgium
- BG-Bulgaria
- CH-Switzerland
- CZ-Czech Republic
- DE-Germany
- DK-Denmark
- EE-Estonia
- ES-Spain
- FI-Finland
- FR-France
- GR-Greece
- HU-Hungary
- IL-Israel
- IT-Italy
- LU-Luxembourg
- NL-Netherlands
- NO-Norway
- PL-Poland
- PT-Portugal
- SE-Sweden
- SK-Slovakia
- UK-United Kingdom

Each NN has its dedicated folder which is then divided into several sub-folders per category.



Divisions into folders in the NN private space on the Specialisation Google drive



Content of a topic or category division in the NN private space. The folder contents the forms, the XLS results and the folders for the uploaded documents (XLS and others)

Figure 4

Each form is divided in 4 mains sections:

- A) The identification of the institution / National Node
- B) The information related to the topic
- C) The personal information of the user and the associated GDPR agreement
- D) The completion status of the document (the user can reedit a form until the information is considered as definitive).

The screenshot shows a form titled "Belgium" with a blue header. The main section is "Institution(s) \*" with a subtext "Select the institution in the list or choose several institutions if you answer as National node." Below this is a list of seven institutions, each with a checkbox: BE-MBG Meise Botanic Garden, BE-RBINS Royal Belgian Institute of Natural Sciences, BE-RMCA Royal Museum for Central Africa, BE-INBO Research Institute for Nature and Forest, BE-KDMA Royal Zoological Society of Antwerp, BE-UNAMUR University of Namur, and BE-VLIZ Flanders Marine Institute. There is also an "Other" checkbox. Below the list is a text input field for "If 'Other', provide the list of institutions represented by the NN answer" with an "Add File" button. At the bottom are "Back", "Next", and "Clear form" buttons.

A) Identification of the Institution

The screenshot shows a form titled "Description" with a blue header. It contains three main sections: "Facility Description" with a text area and "Your answer" label; "Description file (flyer,poster,etc) in PDF" with an "Add File" button; and "Facility Image (jpg, png max 1Mb)" with an "Add File" button. Below these is a "Facility URL" section with a text input field and "Your answer" label. At the bottom is a "Service for external users" section with "Yes" and "No" radio buttons.

B) Information related to the topic

The screenshot shows a form titled "Personal data agreement" with a blue header. The first section is "Who I'm \*" with four radio button options: "I'm the person of contact", "I'm the coordinator at the institution level and I have the authorization of the contact person to fill the information", "I'm the coordinator at the institution level and I don't have the authorization of the contact person", and "Other:" followed by a text input field. The second section is "Agreement \*" with three checkboxes: "The personal data can be used for anonymized statistics and reports", "The personal data can be used for specialized nominative lists of contacts", and "The personal data can not be used for any purpose(s)". At the bottom are "Back", "Next", and "Clear form" buttons.

C) Personnel data and GDPR agreement

The screenshot shows a form titled "Form completion" with a blue header. The main section is "The form state is \*" with three radio button options: "Draft", "Complete", and "Complete and validated. It can be exported to the central repository". Below this is a text line: "A copy of your responses will be emailed to the address that you provided." At the bottom are "Back", "Submit", and "Clear form" buttons.

D) Completion status of the form

Figure 5

## Dashboard of the data encoded in the Google forms or uploaded as XLS files

My Drive > ... > \_Aggregated Data > Statistics ▾

Name ↑	Owner
working	me
Statistics dashboard Forms	me
Statistics dashboard XLS Forms	me

Two spreadsheets / dashboards were developed to report on the completion status of the Specialisation tool. These dashboards provide an analysis of the number of [forms](#) filled in and of the number of [uploads as XLS files](#) for each topic and each NN.

Figure 6

ISO	Country	Exhibitions	Galleries	Facilities	Scientific Instruments	Research fields	Individual Expertises	Institution Expertises	Trainings	Collections lists	Collection details	Total
AT	Austria	0	0	0	0	0	0	0	0	0	0	0
BE	Belgium	0	0	2	2	20	2	10	5	2	3	46
BG	Bulgaria	0	1	2	1	2	0	1	0	1	5	13
CH	Switzerland	0	0	0	0	0	0	0	0	0	0	0
CZ	Czech Republic	1	1	0	0	2	0	1	0	8	8	21
DE	Germany	0	0	0	0	0	0	0	0	0	0	0
DK	Denmark	1	2	1	1	3	9	1	1	3	0	22
EE	Estonia	0	0	0	0	0	1	0	0	1	0	2
ES	Spain	0	0	36	66	46	197	1	0	2	2	350
FI	Finland	0	0	6	3	7	21	0	0	0	0	37
FR	France	6	5	3	2	12	2	2	33	9	18	92
GR	Greece	0	0	0	0	0	0	0	0	0	0	0
HU	Hungary	1	0	1	0	1	2	0	0	0	0	5
IL	Israel	0	0	0	0	0	0	0	0	0	0	0
IT	Italy	0	0	31	11	88	13	0	0	0	0	143
LU	Luxembourg	6	1	1	1	0	1	0	0	0	0	10
NL	Netherlands	0	0	1	1	5	2	0	0	5	2	16
NO	Norway	0	0	0	0	0	0	0	0	1	0	1
PL	Poland	0	0	0	0	0	0	0	0	0	0	0
PT	Portugal	0	0	1	14	0	30	0	0	1	1	47
SE	Sweden	0	0	0	0	0	0	0	0	0	0	0
SK	Slovakia	0	0	0	0	0	0	0	0	0	0	0
UK	United-kingdom	0	0	0	0	0	0	0	0	0	0	0
DISSCo	Total	15	10	85	102	186	280	16	39	33	39	806

Table 1 : Dashboard providing the statistics on the number of forms filled in by the users.

ISO	Country	Exhibitions	Galleries	Facilities	Scientific Instruments	Research fields	Institution Expertises	Trainings	Collections MIDS	Total
AT	Austria	0	0	0	0	0	0	0	0	0
BE	Belgium	0	0	0	0	0	0	0	0	0
BG	Bulgaria	0	0	0	0	0	0	0	0	0
CH	Switzerland	0	0	0	0	0	0	0	0	0
CZ	Czech Republic	0	0	0	0	0	0	0	0	0
DE	Germany	0	0	0	0	0	0	0	0	0
DK	Denmark	0	0	0	0	0	0	0	0	0
EE	Estonia	0	0	0	0	0	0	0	4	4
ES	Spain	0	0	0	0	0	0	0	0	0
FI	Finland	0	0	0	0	0	0	0	0	0
FR	France	1	1	1	2	1	1	0	0	7
GR	Greece	0	0	0	0	0	0	0	0	0
HU	Hungary	0	0	0	0	0	0	0	0	0
IL	Israel	0	0	0	0	0	0	0	0	0
IT	Italy	0	0	0	0	0	0	0	0	0
LU	Luxembourg	0	0	0	0	0	0	0	0	0
NL	Netherlands	0	0	0	0	0	0	0	0	0
NO	Norway	0	0	0	0	0	0	0	0	0
PL	Poland	0	0	0	0	0	0	0	0	0
PT	Portugal	0	0	0	0	0	0	0	0	0
SE	Sweden	0	0	0	0	0	0	0	0	0
SK	Slovakia	0	0	0	0	0	0	0	0	0
UK	United-kingdom	0	0	0	0	0	0	0	0	0
DISSCo	Total	1	1	1	2	1	1	0	4	11

Table 2 : Dashboard providing the statistics on the number of Excel files uploaded by the users.



## 2.2.2. Modifications of the forms using Google scripts

The specialisation exercise will certainly evolve in the future and therefore it was important to anticipate these expected circumstances in the earlier phases of development of the tool. The use of the Google Forms has several advantages, one of which is that it is possible to modify several forms in batch using Google scripts and thus avoiding the individual update of each of the 230 forms.

Still, the large number of forms makes it difficult to adapt them manually. To overcome this constraint a set of **three Google scripts** stored in a specific folder (`_FormsUpdateScripts`) were created. These scripts were developed either from scripts available on the web and created by [Amit Agarwal](#) under MIT License or were created specifically in-house under the direction of RBINS (Patrick Semal main developer) with the help of CETAF (Franck Theeten).

The **first Google script** allows the update of the list of Institutions per country. This script is associated with a google sheet that can be managed by the NN.

B	C	D	E	F
Institution(s)	ISO	Acronym	International name	ROR
BE-INBO Research Institute for Nature and Forest (ROR ID 00j54wy13)	BE	INBO	Research Institute for Nature and Forest	<a href="#">00j54wy13</a>
BE-KDMA Royal Zoological Society of Antwerp (ROR ID 00ed5y156)	BE	KDMA	Royal Zoological Society of Antwerp	<a href="#">00ed5y156</a>
BE-KULeuven Catholic University of Leuven (ROR ID 05f950310)	BE	KULeuven	Catholic University of Leuven	<a href="#">05f950310</a>
BE-MBG Meise Botanic Garden (ROR ID 01h1jbk91)	BE	MBG	Meise Botanic Garden	<a href="#">01h1jbk91</a>
BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22ws83)	BE	RBINS	Royal Belgian Institute of Natural Sciences	<a href="#">02y22ws83</a>
BE-RMCA Royal Museum for Central Africa (ROR ID 001805t51)	BE	RMCA	Royal Museum for Central Africa	<a href="#">001805t51</a>
BE-ULiège University of Liège (ROR ID 00afp2z80)	BE	ULiège	University of Liège	<a href="#">00afp2z80</a>
BE-UNAMUR University of Namur (ROR ID 03d1maw17)	BE	UNAMUR	University of Namur	<a href="#">03d1maw17</a>
BE-VLIZ Flanders Marine Institute (ROR ID 0496vr396)	BE	VLIZ	Flanders Marine Institute	<a href="#">0496vr396</a>

Table 3: List of institutions per country with their ISO, Acronym, Affiliation and ROR. editable thanks to the dedicated Google script.

Each NN can therefore update the list of its contributing institutions belonging to its country by defining the Acronym, the International Name and the ROR of any new institution(s) and then automatically updating all the forms for the country.

**The second Google script** allows the modification of the field names and the help texts of the forms. Several scripts were prepared to modify all the forms under one topic/category (all countries) and all the forms under one country (all topics/categories).

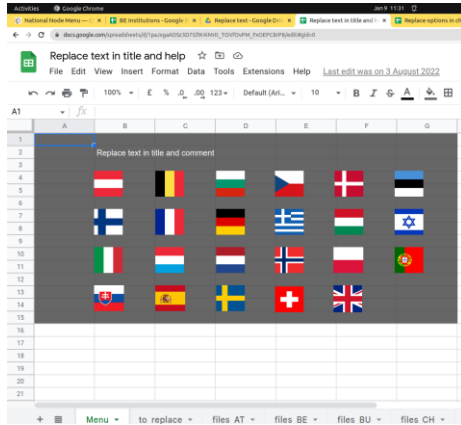


Figure 7

The user has to specify

- the names and comments to be changed in the “to\_replace” sheet.
- the UID of the forms are specified in the files\_NN sheets (already prepared)

The script can be executed using the links that appear on the Menu by clicking on the country flag.

The third Google script can be used to modify the categories proposed in the (multiple)choices field(s)

DISCO	
Distributed System of Scientific Collections	
<p><b>How to update the list of institutions in all forms</b></p> <ol style="list-style-type: none"> <li>1. Modify the name of the field to change in cell B1</li> <li>2. Modify the options list in the column B</li> <li>3. Specify the list of targeted forms in the script</li> <li>4. Click once on the "update forms" button</li> </ol>	
<div style="border: 1px solid black; padding: 5px; display: inline-block;">Update forms</div>	
<b>The form state is</b>	
	Draft
	Complete
	Complete and validated. It can be exported to the central repository

Figure 8

This is important in order to easily update the options proposed for a field and to follow the evolution of the standards such as the inclusion of the [Latimer core](#) for the description of the collections, which is currently under development .

### 2.2.3. Access to the Google forms and to the Google sheets under the NNs spaces

Specific html menus providing direct access to the Google Forms were created for the Institutions and for the NN. They have been integrated in the Plone CMS but any other CMS providing the option with private access to its content can be used for this part of the Specialisation tool. The advantage of Plone CMS is that it is possible to create public pages within private folders. This is not a common feature in CMS which often manage the access rights at the level of the private space and not of the document/object. Thanks to this feature, a direct link can be sent to the institution that will fill the information without any authentication requirement. Nevertheless, the institution Fields menu is located in a private folder of the Plone CMS and is not visible to robots or to anonymous visitors.

	Title	Size	Modified	State
☐	Institutions	1 KB	May 17, 2022 01:46 PM	Published
☐	Collections	1 KB	Dec 21, 2020 05:38 PM	Published
☐	belgium-flag-square-icon-128.png	1.1 KB	Dec 21, 2020 05:38 PM	Published
☐	Specialisation	1 KB	Aug 03, 2022 11:07 AM	Private

Figure 9: Belgium private space in the Plone CMS. The Specialisation folder with the menus is private.

**Institution Menu**  
by mansadmin — last modified Jun 17, 2022 10:28 AM — History

You need to fill a new form for each piece of information you want to add (e.g. each Research Field is encoded individually)

**Each Form is divided in 3 sections:**

- The first one allows you to identify the institution or the person.
- The second one asks information about the topic (Research Field, Individual expertise, ...)
- The last one is related to the GDPR regulation and the state of the form

A link will be automatically send to the email address provided in the first section. It allows to re-edit the form if some information was still missing.

**When the form is complete.**  
Choose "Complete and validated. It can be exported to the central repository" for the state of the form.

Topic	Form (*)
 Collection(s) list	
 Collection(s) full information	
 Research Fields	
 Individual Expertises	
 Institution Expertises	
 Facilities	
 Scientific Instruments	
 Galleries	
 Exhibitions	
 Trainings	

**DiSSCo Specialisation Plan User manuals**

- Collection(s) list
- Collection(s) Full Information
- Research Fields
- Facilities
- Scientific Instruments
- Institution Expertises
- Individual Expertise
- Trainings
- Galleries
- Exhibitions

Figure 10: The Institution menu in Plone CMS presenting the topics/categories and the links to the forms as well as the links to the User manuals for each topic/category.

In each Plone folder of the Specialisation tool corresponding to a country, the "Institution menu" is public, and is the default view of the folder. It does not need authentication from the user receiving the link from the NN Coordinator.

Select: All				
	Title	Size	Modified	State
☐	National Node Menu	15.4 KB	Aug 04, 2022 10:22 AM	Private
☐	Institution Menu	23.2 KB	Aug 01, 2022 12:03 PM	Published

Figure 11: The Status of the NN menu folder is private but the Institution menu is set as “Published” making it accessible through a direct link.

A second menu has been developed specifically for each NN.

### National Node Menu

by marsadmin — last modified Jan 15, 2023 09:35 AM — [History](#)

This menu help you to follow the encoding by the institutions in your country.



- [EDIT the Institutions list](#)
- [View the Institution menu \(used to fill the information\)](#)

See the surveys encoded for your country :

Topic (click on topic to open the google folder)	Responses XLS form	Uploaded XLS files
<ul style="list-style-type: none"> <li>• <a href="#">Collections list</a></li> <li>• <a href="#">Collections Full Information</a></li> <li>• <a href="#">Collections MIDS</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Research Fields</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Individual Expertises</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Institution Expertises</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Facilities</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Scientific Instruments</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Galleries</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Exhibitions</a></li> </ul>		
<ul style="list-style-type: none"> <li>• <a href="#">Trainings</a></li> </ul>		

Figure 12: The NN menu with the links to the spreadsheets collecting the results and to the uploaded XLS files.

The link to this private “NN menu” is available by clicking on the flag of the country in the public “Institution menu” of each country.

Each NN of DiSSCo has received from the “Specialisation tool” administrator a NN Coordinator account with user name and password giving them access to this private menu.

This menu gives access to:

- the Google Drive sub-folders of the country and to the files that the institutions will upload under the different categories/topics as well as the aggregated data related to one Google form.
- the uploaded XLS files

The NN Coordinator is provided with one link to the list of the institutions that are members of DiSSCo in the country and another link to access to the Institution Menu.

## 2.2.4. Aggregation of the data using Google scripts

All the data are recorded at the NN level. This has the advantage of providing an access to the information of the NN but can be a limitation if the Specialisation analysis has to be done at the Infrastructure level.

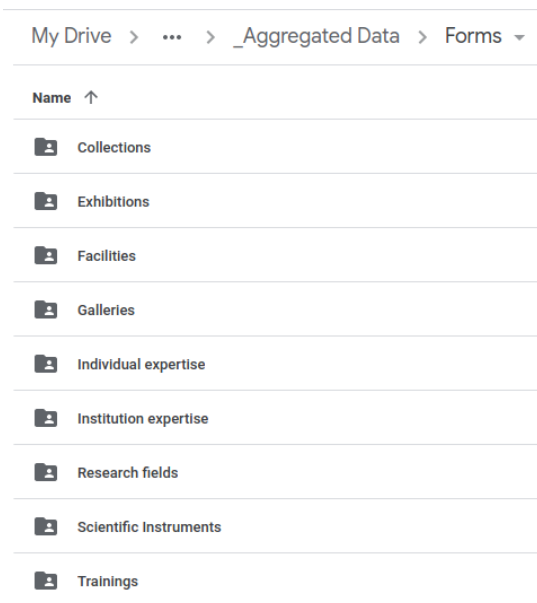


Figure 13: DiSSCo aggregate data space divided in topic sub-folders

In order to address this, a folder with aggregated data (\_Aggregated Data) was created in the Google space. Thanks to scripts developed to aggregate the NN google sheets at the level of the DiSSCo infrastructure and able to merge the data collected using XLS files an analysis at infrastructure level is made possible.

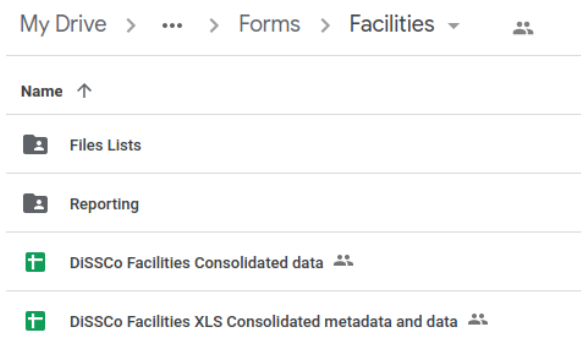


Figure 14: Topic sub-folders with the XLS aggregated files and the folders related to the list of the files and the reporting of the data with automatically prepared figures

For each category/topic, the subfolder is composed of the XLS file aggregating the information of the forms and of the uploaded XLS files.

A specific Reporting folder contents several files prepared to display the results as a dashboard at the national level.

Scripts were created specifically to aggregate data. Each script is directly associated with the corresponding aggregated google sheet.

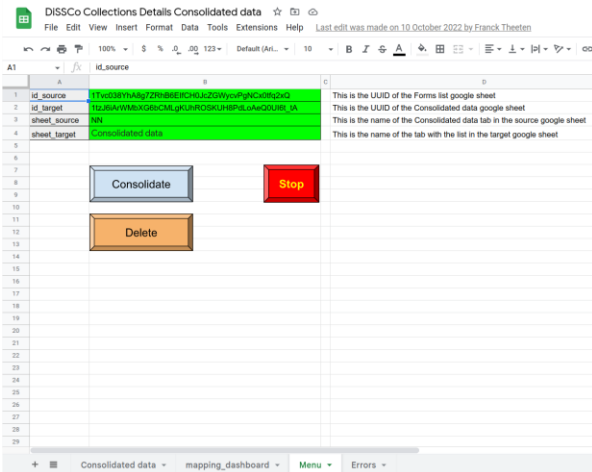


Figure 15: Menu of the consolidation script

A menu tab allows the user to specify:

- the id\_source UUID with the UUIDs of the NN data results Google sheets which have to be merged. This is already prepared but can be modified by the user.
- the id\_target which is the UUID of the Consolidated data google sheet

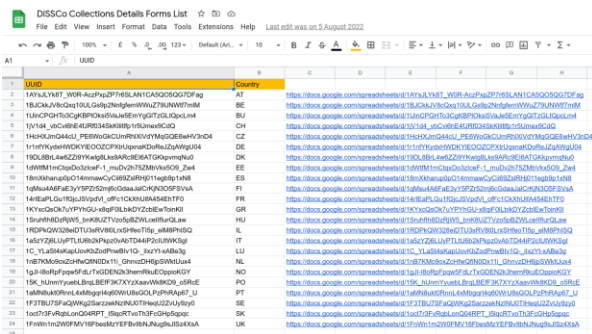


Figure 16: List of Sheets to consolidate

For each topic, a list of the Google forms having to be merged is prepared with the identification of the form and the direct link to access to the form.

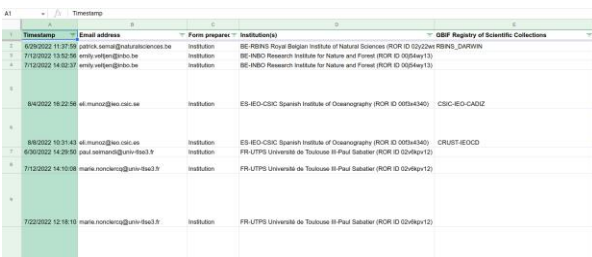


Figure 17: Consolidated sheet

All data is aggregated in the Consolidated data tab. These data will be exported to ElasticSearch indexes using web services.

## 2.2.5. Creation of graphs and maps for the analysis of the data

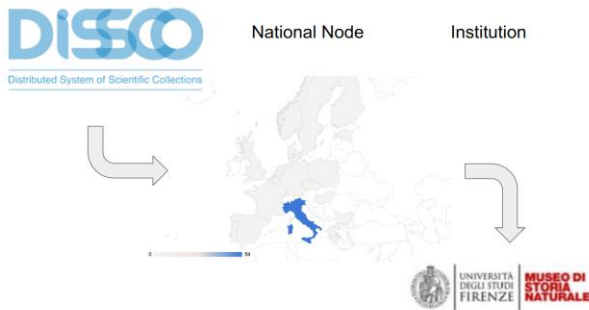


Figure 18

The aim of the Specialisation tool is to search and analyse the data provided by the institutions and Nodes at the institution, NN and DiSSCo levels through the topics/categories (collections, facilities, scientific Instruments, exhibitions, galleries, expertises and training) to extract the assets of the Community.

Several Google sheets were created for analysis using different predefined figures.

The screenshot shows a Google Sheet with the following data:

Country	Institution(s)	Title of the Research Field (in English)	Abstract
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Protected Nature	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Soil & air	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Forest	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Wildlife management	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Invasive species	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Climate	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Agriculture	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Nature & society	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Water	
Belgium	BE-INBO Research Institute for Nature and Forest (ROR ID 0054wy13)	Data & infrastructure	
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	EVOLUTION, THE DRIVING FORCE OF LIFE	Evolutionary proce generations to bil
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	BIODIVERSITY INVENTORY	Under this Researc
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	ECOSYSTEMS DYNAMICS	In this Research 1
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	PAST INTERACTIONS BETWEEN HUMANS AND	The relationships humankind gradual now speak of huma
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	BLUE GROWTH AND MARINE MANAGEMENT	Seas and particul but are greatly ur is indispensable.
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	BIODIVERSITY AND NATURAL RESOURCES PE	One of the RBINS'
Belgium	BE-RBINS Royal Belgian Institute of Natural Sciences (ROR ID 02y22w83)	SUSTAINABLE USE OF GEO-RESOURCES	It supplies credit Our need for natur natural resources

Figure 19: Analysis spreadsheet

Each analysis Google sheet is divided into several tabs:

- The **Results data** tab which is a dynamic copy of the Aggregated data file using the “=Research Fields Results data!D2” function
- The **Cleaned data** which is a cleaned version of Results data and used for the Analysis
- The **Statistics** which displays the number of forms used
- The **Analysis** tab with all the pre designed figures
- The **Collections** tab with the figures related to the collections

The produced figures can be **graphs**:

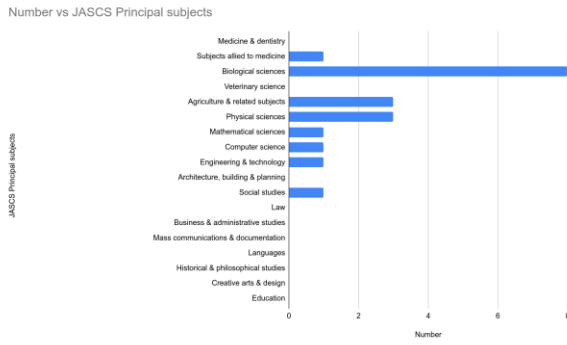


Figure 20: Correlation between the research field and the JASCS subjects

**Digitisation instrument**

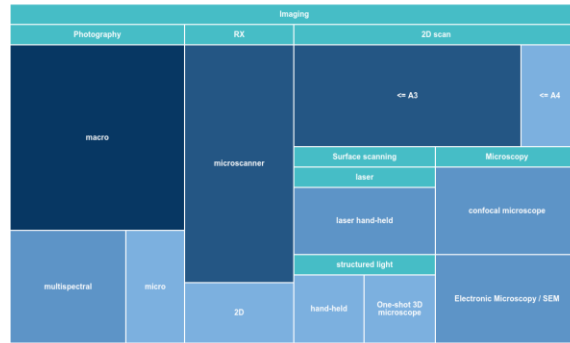


Figure 21: Distribution of digitization instruments

The produced figures can also be **maps**:

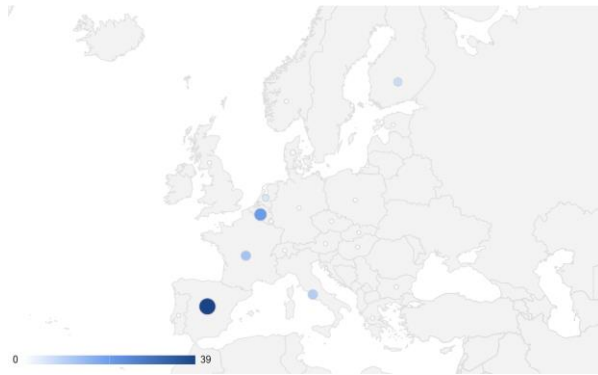


Figure 22: Number of answers by NN

Maps are produced through the out of the box Google Sheet application.

They help visualise the distribution of data per country. However, it is not possible to plot specific locations using longitude and latitude displaying distribution maps at the institution level with a link to the specific information.

This requires the use of Google Map / Earth, which are costly products.

We developed a specific service which produces maps with points and hyperlinks using data encoded in google sheets. This is a generic tool that can serve other purposes.



 Geographic map

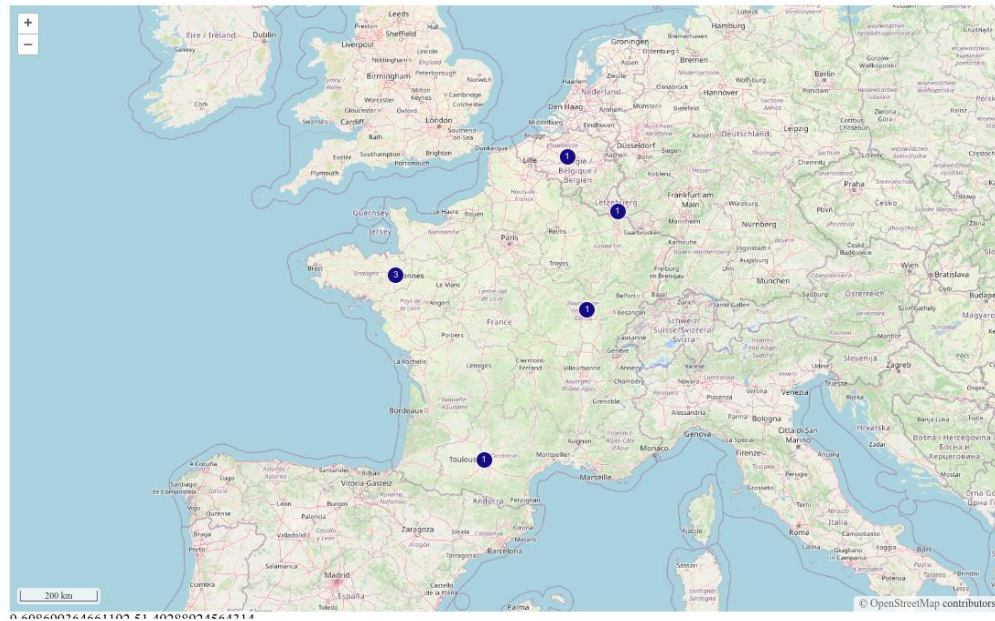


Figure 23: Number of answers by Institution

The `test_url_openlayers` google sheet is composed of 5 columns (*key*, *pt*, *col\_def*, *size*, *urls*) defining a specific map for each line. The map is generated using the Open Street Map background (Open Source).

<b>Column</b>	<b>Example</b>
The <b>key</b> value defines the map_id	Map-1
The <b>pt</b> value defines the position of the points using the following syntax:  The syntax has to be repeated to display several points	<b>point_id: long;lat;point_label </b>  1:4.37669555963812;50.8366740498444;RBI NS
The <b>col_def</b> defines the color of the symbol	76ff56
The <b>size</b> defines the size of the symbol	15
The <b>urls</b> defines the hyperlinks related to the points using the following syntax:  The syntax has to be repeated to match to the <b>pt</b> list of points	<b>point_id: url</b>  1:https://www.naturalsciences.be

Table 4: The syntax of the Map sheet

The **pt** values can directly be imported from the Google aggregated data using the IMPORTRANGE function e.g.

```
=IMPORTRANGE("https://docs.google.com/spreadsheets/d/19_LgADGoKsNhbO3z7WQiXz0nZA3GVPQWlyRiqUGbuoI/", "MAPS DATA!J1")
```

The resulting **pt** line is

```
"1:4.37670673072384;50.8366773797725;1-BE-RBINS Royal Belgian Institute of Natural Sciences|2:-1.63984167964406;48.1197512323933;2-FR-UR1 University of Rennes 1 (ROR ID 015m7wh34)|3:-1.6396315569331;48.1200317773575;3-FR-UR1 University of Rennes 1 (ROR ID 015m7wh34)|4:1.449444;43.594167;4-FR-MHNT Muséum d'histoire naturelle de Toulouse (ROR ID na)|5:-1.6401568637105;48.1195408226647;5-FR-UR1 University of Rennes 1 (ROR ID 015m7wh34)|6:5.06532089826108;47.316265904582;6-FR-UB University of Burgundy (ROR ID 03k1bsr36)|7:6.1357;49.6098;7-LU-MNHNL The National Museum of Natural History (ROR ID 05natt857)"
```

and the **urls** is

```
7:https://www.mnhn.lu/de/dauerausstellungen/|6:https://google.com|5:https://culture.univ-rennes1.fr/instruments-scientifiques|4:https://www.museum.toulouse.fr/l-exposition-permanente|3:https://culture.univ-rennes1.fr/geologie|2:https://culture.univ-
```

rennes1.fr/zoologie|1:https://www.naturalsciences.be/en/museum/exhibitions-view/771/2762/697|

The map view can be directly viewed in a web page e.g.

[https://darwin.naturalsciences.be/ol\\_service/index\\_gs.php?key=Map-5](https://darwin.naturalsciences.be/ol_service/index_gs.php?key=Map-5)

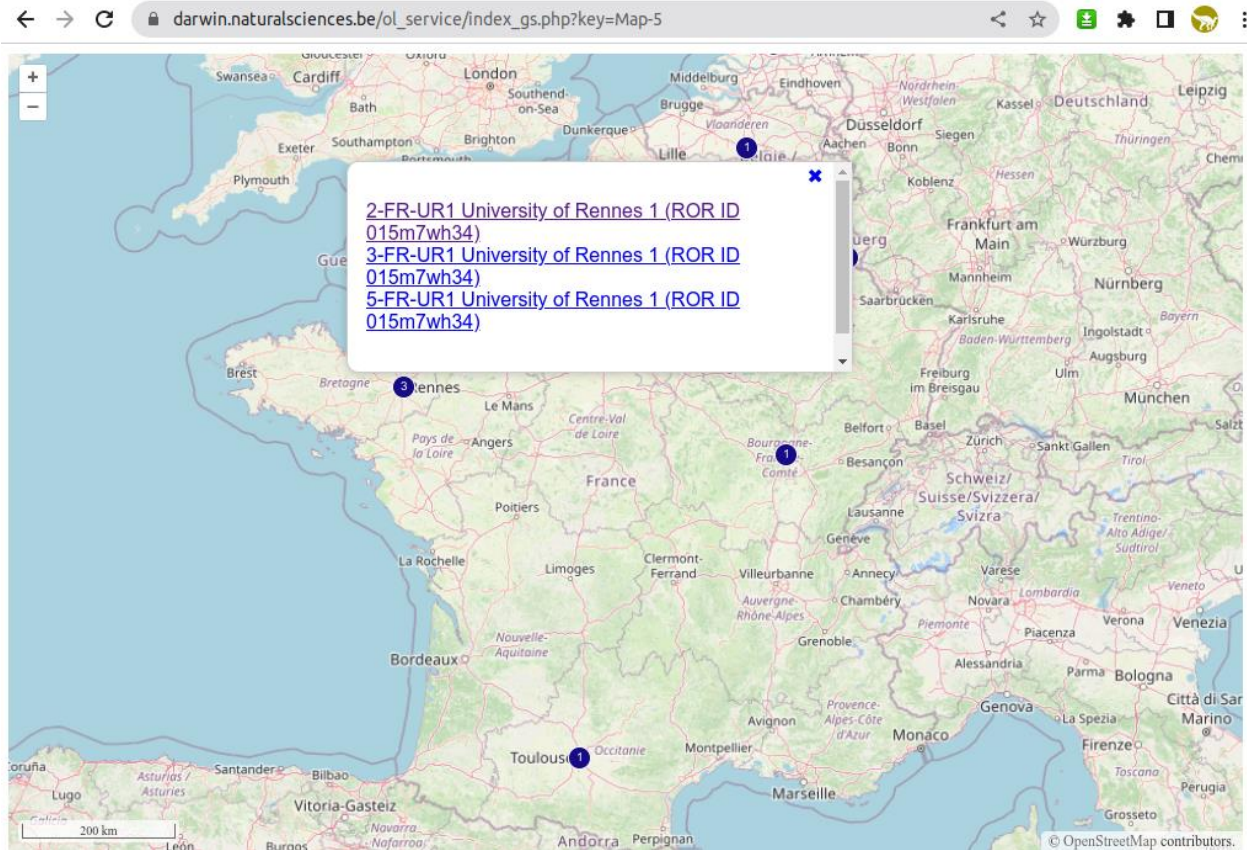


Figure 24: Maps with individual location and links

or integrated in a specific Institution / NN / CETAF / DiSSCo web site using Iframes.

All figures produced by the Analysis Google sheets can be reused in a **dynamic dashboard** using the "Publish to the web" option:

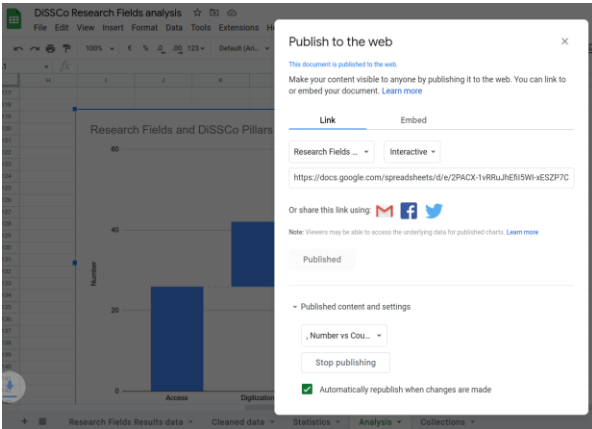


Figure 25: Publishing the figures

This option creates a link to the figure and the code to embed the figure in an external page. It allows to define:

- the state of publication of the figure
- if automatically update the figure in the dashboard when the data are updated in the Google sheets.

In this prototype, we embed the figures in dedicated pages of the Plone CMS using the Iframe tool.

This allows the user to create a flexible dashboard “on demand” which can be easily updated following the specific user requirements in the next steps of the development.

## Reporting

by marsadmin — last modified Aug 22, 2022 03:52 PM — [History](#)



Figure 26: The reporting menu

## Geographic distribution map of the answers

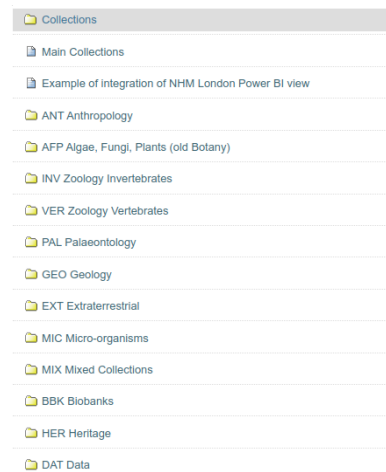
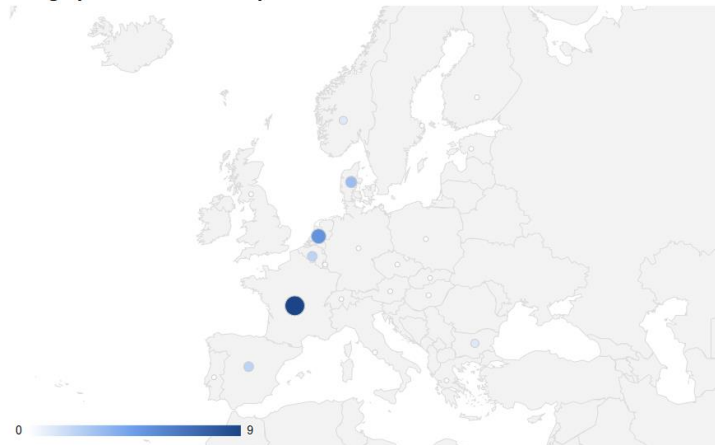
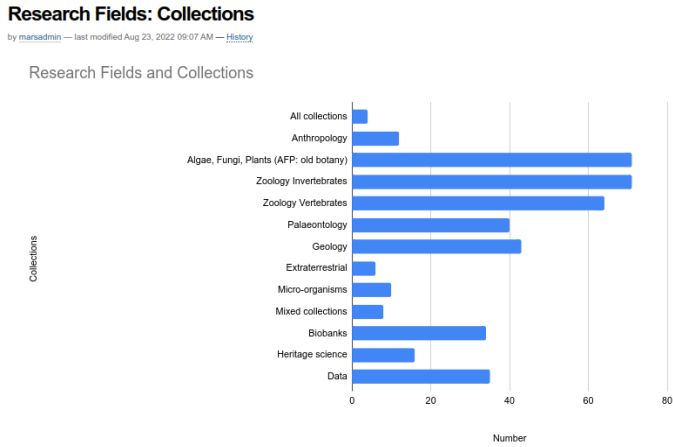


Figure 27: Distribution of the answer

Figure 28: Link to the Collections views



A figure displaying the correlation between the topic fields and the collection and/or the sub-collection is prepared and available when applicable.

Figure 29: Correlation between research field and Collections

The **iframe approach** also allows the inclusion of external tools in the reporting template as the Collections Digitization dashboard developed by the NHM.

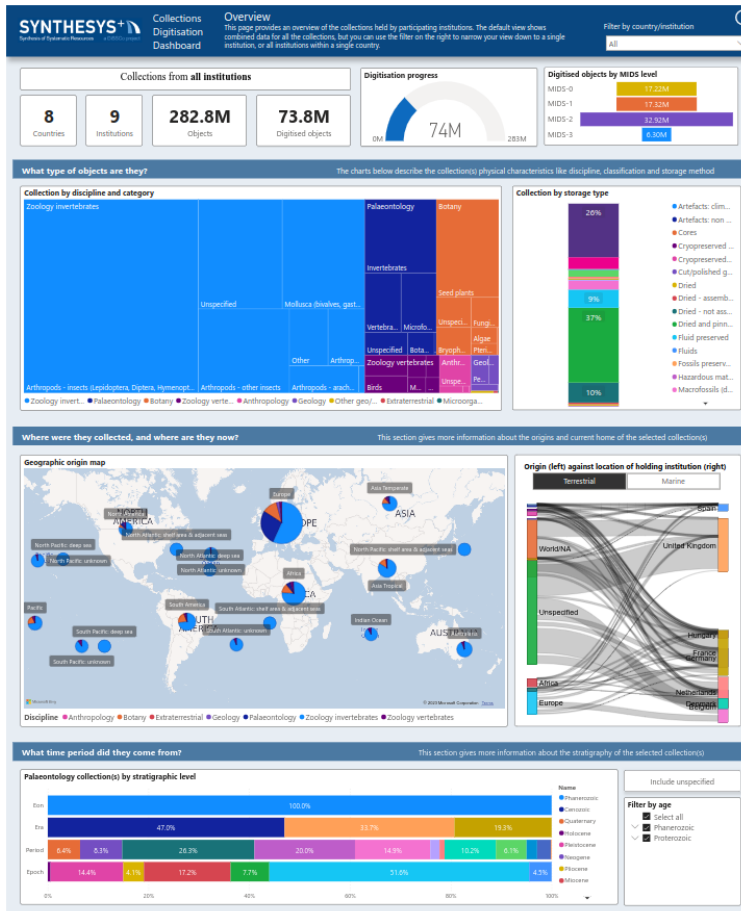


Figure 30: Synthesys + Collection Digitization Dashboard

## 2.2.6. Data analysis and search

Google sheets present many advantages when it comes to data analysis through figures and predefined filters but this tool has also one limitation:

- It is not easy to create a global search interface with multiple criteria which can be used by the NN or at the DiSSCo level for a complex query e.g.:

“Which institution(s) hold a microfossils collection curated by a scientist expert in African specimens and where a Scanning Electron Microscope is available?”

Such a complex query with “open” questions should be based on external indexes as ElasticSearch indexes. These indexes can also be used as an authority for the external services as DiSSCo ELViS or the CETAF website. We have addressed such an issue and have developed a formula to solve this constraint by indexing the data in ElasticSearch Indexes.

## 2.2.7. Indexation of the data in ElasticSearch indexes and access by external users



ElasticSearch indexes were specifically developed to process the information on the institutions, collections, facilities and expertise (individual). They are already used to feed the CETAF website based on a Wordpress CMS. More indexes will be developed for the other three remaining categories namely exhibitions, galleries and training.

The import of data from Google sheets to ElasticSearch Indexes is achieved using Python scripts connecting the Google API and filling the indexes. The scripts also compute several values as MIDS derived from raw data filled in the forms standardising the subdivision of the collections following the controlled vocabularies.

The resulting ElasticSearch indexes are accessible through specific access point by external services using dedicated code in Ajax or similar technologies:

- [https://darwin.naturalsciences.be/elasticsearch\\_cetaf\\_passport/institutions/](https://darwin.naturalsciences.be/elasticsearch_cetaf_passport/institutions/)
- [https://darwin.naturalsciences.be/elasticsearch\\_cetaf\\_passport/collections/](https://darwin.naturalsciences.be/elasticsearch_cetaf_passport/collections/)
- [https://darwin.naturalsciences.be/elasticsearch\\_cetaf\\_passport/facilities/](https://darwin.naturalsciences.be/elasticsearch_cetaf_passport/facilities/)
- [https://darwin.naturalsciences.be/elasticsearch\\_cetaf\\_passport/expertises/](https://darwin.naturalsciences.be/elasticsearch_cetaf_passport/expertises/)

Each dataset can be accessed in JSON format via ElasticSearch

```
JSON Raw Data Headers
Save Copy Collapse All Expand All (slow) Filter JSON
{
  "_index": "cetaf_passport_collections",
  "_type": "_doc",
  "_id": "BE-MBG_BOTANY",
  "_score": 1,
  "_source": {
    "to_parent_institution": "BE-MBG",
    "institution_name": "MBG Meise",
    "coverage_fields": {
      "main_category": "Botany",
      "discipline": {
        "0": "Algae",
        "1": "Bryophytes",
        "2": "Pteridophytes",
        "3": "Seed plants",
        "4": "Fungi_Lichens (Including Myxomycetes)",
        "5": "Other"
      }
    },
    "sub_collections": {
      "0": {
        "size_and_digitisation_fields": {
          "specimens_count": "153",
          "discipline": "Algae"
        }
      },
      "1": {
        "size_and_digitisation_fields": {
          "specimens_count": "400",
          "discipline": "Bryophytes"
        }
      },
      "2": {
        "size_and_digitisation_fields": {
          "specimens_count": "80000",
          "discipline": "Pteridophytes"
        }
      },
      "3": {
        "size_and_digitisation_fields": {
          "specimens_count": "2932",
          "discipline": "Seed plants"
        }
      }
    }
  }
}
```

Figure 31: JSON display from the ElasticSearch data

The screenshot shows a REST client interface with a 'History' panel on the left and a 'Request' panel on the right. The 'Request' panel shows a GET request to `https://collections.naturalsciences.be/cpb/nh-collections/countries/germany/de-zfmk/2-cetaf-passport-collections/collections/vz`. The response is a 200 OK status with a JSON body containing the index mapping for `"cetaf_passport_institutions_full"`.

```
{
  "cetaf_passport_institutions_full": {
    "aliases": {},
    "mappings": {
      "properties": {
        "identification_fields": {
          "properties": {
            "non": {
              "type": "text",
              "fields": {
                "keyword": {
                  "type": "keyword",
                  "ignore_above": 256
                }
              }
            }
          }
        }
      }
    }
  }
}
```

Figure 32: Access from the ElasticSearch indexes via a REST client (web service)



A prototype of the search interface based on the interface of the CETAF website has been created. This is the proof of concept of the technical approach. It needs to be further developed to be used by the NN and DiSSCo RI.

The screenshot shows a search interface with the following components:

- Filters:** Five dropdown menus for "Select an institution", "Select main category", "Select a discipline", "Select an area", and "Select country". A "Reset" button is located below the filters.
- Show:** A dropdown menu set to "10" with the text "entries" below it.
- Table:** A table with columns: Main Category, Discipline, Geographical Areas, Country, Specimen Count, and Institution.
 

Main Category	Discipline	Geographical Areas	Country	Specimen Count	Institution
Anthropology	Unspecified Human Biology Archaeology Other			0	MFN-Berlin
Anthropology	all			0	Royal Belgian Institute of Natural Sciences
Anthropology	all			23,000	CSIC_MNCN
Anthropology	Human Biology	Terrestrial Africa Terrestrial Asia Temperate		39,500	HNHM Hungary

Figure 33: Prototype of the GUI for the multi-criteria search interface derived from the CETAF website search interface.

## Workflow of the information

The first version of the Specialisation tool proposed the following workflow:

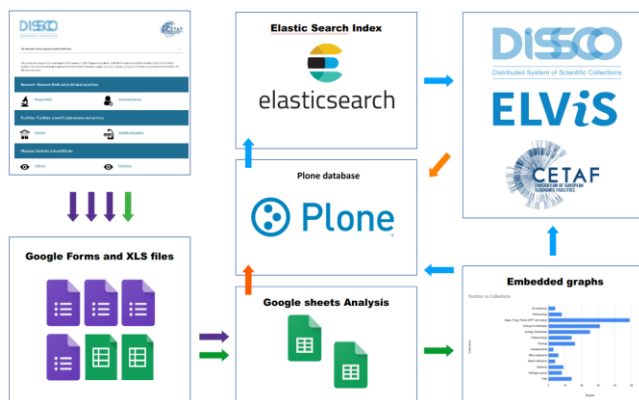


Figure 34: Workflow of the version 1 with Plone as centralised database

1. The data are encoded via google forms and google sheet files starting from a google site menu. Specific access to the forms have to be provided.
2. The data are aggregate manually and exported to the centralised Plone CMS
3. The data of the Plone CMS are exported to several ElasticSearch indexes using API
4. The data are used to produce a graphic dashboard using google sheets
5. The ElasticSearch indexes are used as data authority by the external services as the DiSSCo ELViS or the CETAF website

6. The figures of the dashboard are included in the reporting system using embedded graphs or Iframes

The second version of the Specialisation tool proposed a simplified workflow:

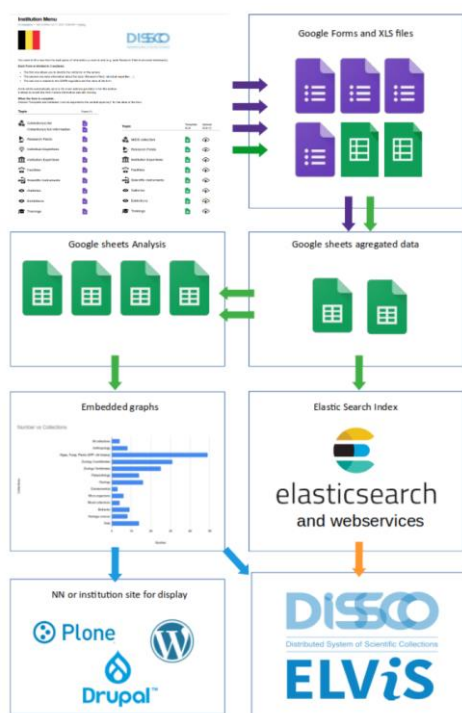


Figure 35: Workflow of the version 2 with ElasticSearch as centralised indexes

1. The data are encoded via google forms and google sheets starting from a Plone site menu. This allows the use of direct links and a private folder for security reasons.
2. The data are aggregated automatically using google scripts
3. The data are exported directly from the aggregated google sheets to several ElasticSearch indexes using API. Plone is not used anymore as a centralised database.
4. The data are used to produce a google sheets analysis and figures for a dashboard
5. The ElasticSearch indexes are used as data authority by the external services as the DiSSCo ELViS or the CETAF website
6. The figures of the dashboard are included in the reporting system using embedded graphs or Iframes.

This reporting system can be developed with any CMS or even included directly in the DiSSCo website if needed.

### 2.2.8. Update of the information

The collection and the update of the information could be a time consuming task for DiSSCo partners. Through Google forms, pre-filled forms can be sent which have just to be updated with the latest values. This is a way to limit the task of encoding. It also helps the distribution of the work among the staff members as a simple link is enough to update the information without the creation of hundreds of different user accounts.

The frequency of updates has to be defined as to ensure global, harmonised and coherent visualisation of data..

For each update period, an email will be sent to the address registered in the prefilled form with a link to the form ready for any updates. The link can also be forwarded to the new person in charge of the encoding if the previous one left.

1	Google Form Link	Timestamp	Institution(s)	Title of the expertise (in English)	Email address	Abstract (in English)	Do you want to complete
2	<a href="https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1">https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1</a>	16/08/2022 15:42:55	BE-INBO Research Institute	Collection management	emily.veljen@inbo.be		No, I want to use the OR
3	<a href="https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1">https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1</a>	16/08/2022 15:43:07	BE-RBINS Royal Belgian Institute of Natural Sciences	Q2	patrick.semal@naturalsciences.be		Yes, I want to complete t
4	<a href="https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1">https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1</a>	17/08/2022 08:45:27	BE-RBINS Royal Belgian Institute of Natural Sciences	Anthropology	patrick.semal@naturalsciences.be		Yes
5	<a href="https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1">https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1</a>	17/08/2022 10:40:51	Value 1	Value A	patrick.semal@naturalsciences.be	Test list 4	Yes
6	<a href="https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1">https://docs.google.com/forms/d/e/1FAIpQLSeGf4ORjlaeU-rj8_xRccduLvXemvejeIJNmSd1</a>	17/08/2022 10:55:16	Value 3	Value A	patrick.semal@naturalsciences.be	Q3	Yes

Figure 36: The link to the prefilled form is automatically created from the aggregated data google sheet

To ensure that the update is not limited exclusively to the information encoded in the aggregated data Google sheets, some further improvements need to be made during next phases of DiSSCo RI development.

A possible solution could be to dynamically create the prefilled Google Form using a link and the syntax of the Google form but edited directly from the ElasticSearch indexes instead of the Google sheets. This allows to incorporate in the prefilled Google Form some data directly imported in the ElasticSearch Indexes from previous exercises (e.g. Synthesys+ collection dashboard).



Figure 37: Update workflow of the data using Google and python scripts

## 2.3 Knowledge base Integration – ( links with existing dashboards )

One of the main perspectives for this tool is to see its manuals and documentation integrated in the Knowledge Base of DiSSCo through a specific landing page. This is a line for further improvement since currently an account in the system in order to add content cannot be created yet, while reaching out to the administrator of the Knowledge Base requires authentication with a user account. This issue will be addressed through the DiSSCo Technical team in order to overcome the bottleneck created around new users' accounts and authentication processes.

DiSSCo Knowledgebase

### User Registration

If you've never logged on to DSpace before, please enter your e-mail address in the box below and click "Register".

E-mail Address:

If you or your department are interested in registering with DSpace, please contact the DSpace site administrators.

[Leave a message for the DiSSCo Knowledgebase administrators.](#)

Figure 38: Knowledgebase: the registration page is not working

## 2.4 Software code

The Specialisation tool is mainly developed with Google forms and Google sheets documents.

The software code is limited to

- the Google Scripts used for the modification of the forms, the aggregation of the data. The App Script code (the script language of Google based on javascript) is associated with all the related forms. The UUID used in Google Scripts provides direct access to the documents and files most of the time without any authentication. For security reasons, the code of the scripts is only available on demand.
- The code is also composed of the Python scripts creating the ElasticSearch Indexes on the base of the Google Forms. The UUID of the Google Sheets are hardcoded in the scripts. For security reasons, the code of the scripts is only available on demand.
- The code for the generation of maps with points and hyperlinks was developed using PHP / javascript / Open Layers. The PHP code uses the Google Authentication API requiring the

download of a static file with the connection parameters. For security reasons, the code of the scripts is only available on demand.

- The menu for institution and NN as well as the results pages are created using the Plone CMS. No specific code was developed as the Plone CMS is using simple html code using WYSIWIG editor. The creation of the structure and the hierarchy is also managed with a WYSIWIG editor. All features for the definition of the status (public/private) are out of the box. The only code which has to be used is the html code for the use of the Iframe:

## 3. Specialisation highlights

### 3.1 Data collection methodology

The partners to Task 8.1 have been updated on the progress of the development of the tool through several meetings organised at task level throughout the project lifetime. They were presented for discussion with the categories and fields to be covered and their access rights and codes, and the various options for data entry through the Google Forms and then through the upload of their own databases mostly under excel files. The partners were the first user testers of the tool. The tool was also presented and explained at both All Hands Meetings (AHMs) organised within DPP to the other partners of the project, but also on several occasions during the NN meetings, where National Nodes representatives were given a short training on the access of the tool, the goals and expected outcomes, the levels (at National Node or Institution level) at which the data can be entered, the possibilities of data extraction and further analysis, and the data updating and sharing.

At month 26 of the lifetime of DPP, the leader of the exercise submitted Milestone MS8.5 “Initial Findings for the Thematic Specialisation Plan” with which the first steps of the data collection process and tool development were reported. This period was followed by further interactions held between March and August 2022 to reach out to the National Nodes representatives. Personalised emails were sent to the 21 National Nodes leaders members of the DISSCo network<sup>3</sup>. They were given their respective access links to their NN webpage on the Specialisation tool, their respective menu, as well as a presentation as tutorial to help them access the various tool pages and forms.

Each NN was asked to have a bilateral meeting as follow up by the task leader.

As a result of the overall communication campaign towards National Nodes, and including the NN that were also partners in Task 8.1, responses were given from 10 out of the 21 NN (AT, BE, CZ, DE, DK, EE, ES, FI, FR, NL).

Because of the vast and varied amount of information that the tool can collect under each of the categories to be covered (Research fields, Training, Scientific instruments, Individual expertise, Galleries, Facility) and due to time constraints, it was proposed to the NN Coordinators to limit the information requested to their partner institutions to 3 main topics - Research fields, Scientific

---

<sup>3</sup> See the NN Coordinators under each Country link from the DiSSCo.eu website [Network | DiSSCo](#)

instruments, and individual Expertise. Such a focused procedure would allow to obtain some indicative results that could later on be scaled up and enlarged.

More exchanges took place by email and several other multilateral meetings also in the framework of CETAF (such as presentations given to the General Assembly of CETAF). These meetings were additional opportunities to provide training on the tool, collect any issues and provide quick solutions to finetune the tool according to the difficulties encountered. Among the issues found were: how to enter the expertise not related to collections ex from institutions such as universities? What are MIDS? Terminology was harmonised and categorised as to have a common ground and rest of major issues were addressed gradually as the subsequent iterations of the development of the tool took place...

As mentioned before, 10 out of 21 possible responses, from Nodes national level, were obtained. Each Node selected those categories that were considered either more accessible or more relevant to their respective Node's participants. Jointly, the largest set of data compiled were on "Individual expertise" and "Research fields" as these had been set as the main topics to focus at first stage. These came from Nodes that had a well established national network with relatively well known partner institutions and, as Figure 22 illustrates, resulted from the National Nodes that were part of Task 8.1 partnership as it is the case for the following countries: BE, ES, FI, FR and IT.

For the rest of the categories, such as for the list of Scientific Instruments or for the MIDS table the data compiled is scattered. Therefore, for this current exercise the information has not been taken into consideration to avoid providing unreliable results based on partial, fragmented and unbalanced data .

The difficulty to reach out to the partner institutions within a Node should not be underestimated. These partners to the national node need to have a clear understanding of the benefit they can have to work on this process to provide their data on regular bases, and the added value of contributing to yet another platform when they are already providing some of this information to other databases and/or aggregators. Special emphasis should be made to underline the value proposition of having a comprehensive and sustainable unique source of information (through the CETAF Collections Registry towards the DiSSCo RI) that could further be linked to as many other platforms as necessary, while still keeping a sole curation point close to the resources (at institutional level).

The promotion of the Specialisation tool through the NN meetings has been a real help to disseminate and promote the use of this mechanism, and also to improve it from the feedback received. The development achieved under the DPP project is a pilot exercise that requires an extensive additional effort to complete the database with the necessary information that needs to be completed and updated.

### **3.2 First outputs from current exercise**

The Specialisation tool is developed for collecting a (very) large set of data in a distributed way.

Depending on the user's purpose and objectives the request applicable to the data will provide information per country (National Node), institution, institution departments and finally, scientists. The institution ID is the primary key allowing to cross-reference the data from the different forms.

As explained in Chapter 3, the data currently provided constitute a baseline based on the participation of a group of Nodes and institutions. The datasets are good enough to test and evaluate how the Specialisation tool works and help in the setup of the prototype. Still the data are not comprehensive enough to provide definitive results to be used in the setup of the strategy of specialisation of the DiSSCo infrastructure.

In the future according to the management of DiSSCo RI, the new CETAF website and the needs raised from the users, it will be necessary to gather more datasets to enlarge and enrich the picture of the specialisations existing among the community.

The expected specialisation outcomes will provide a complete overview of areas of excellence of the community on top of which services can be built:











1. the collections and sub-collections housed by the DiSSCo institutions and the specialisation hold across institutions and countries;
2. the digitization strategies and set-ups/workflows among the DiSSCo community and the synchronisation with specialised tools such as:
  - a. the DIGIT Key developed in the framework of the Synthesys+ project,
  - b. the [DiSSCo Digitisation Guides](#) developed in the framework of DiSSCo Prepare,
  - c. the digitization dashboard developed in the framework of the Synthesys+ project
3. the research fields driving research in the DiSSCo institutions with the specialisation of the research topics (e.g. marine management or biodiversity inventory);
4. the facilities and scientific instruments available for scientific research or digitization in the DiSSCo infrastructure and the availability for users in the frames of Virtual Access (VA) and/or Transnational Access (TA) through the ELViS system;
5. the expertise existing at the institutional s or at the individual levels, which can be embedded into other products and initiatives such as the Red List of Taxonomists and the CETAF-DEST (Distributed European School of Taxonomy)), or EU funded projects as TETTRIs (Transforming European Taxonomy through Training, Research and Innovations);
6. the available training offer among institutions and the needs of the community for tooling up and provide capacity building to their staffs;
7. the permanent galleries which are one of the important the link with the citizens; and
8. the capacity of the institution to develop touring exhibitions and to propose to other institutions as a multidisciplinary and transnational additional service.

## 4. User Manual

In order to help the tool back-end manager as well as the front-end user to more effectively navigate and meaningfully use the tool, a User Manual has been made available under the form of didactic forms for each of the categories covered.

Since the tool can be used by people with different backgrounds and profiles, each form of the User Manual is dedicated to specific tasks of a certain specific user's characteristics.

The User Manual including its forms were prepared using Google docs allowing collaborative editing, and are visible directly in the Plone Menu Interface from the list of links. See screen below.

DiSSCo Specialisation Plan User manuals	
 Collection(s) list	<ul style="list-style-type: none"><li>● <a href="#">Collection(s) list</a></li></ul>
 Collection(s) Full Information	<ul style="list-style-type: none"><li>● <a href="#">Collection(s) Full Information</a></li></ul>
 Research Fields	<ul style="list-style-type: none"><li>● <a href="#">Research Fields</a></li></ul>
 Facilities	<ul style="list-style-type: none"><li>● <a href="#">Facilities</a></li></ul>
 Scientific Instruments	<ul style="list-style-type: none"><li>● <a href="#">Scientific Instruments</a></li></ul>
 Institution Expertises	<ul style="list-style-type: none"><li>● <a href="#">Institution Expertises</a></li></ul>
 Individual Expertise	<ul style="list-style-type: none"><li>● <a href="#">Individual Expertise</a></li></ul>
 Trainings	<ul style="list-style-type: none"><li>● <a href="#">Trainings</a></li></ul>
 Galleries	<ul style="list-style-type: none"><li>● <a href="#">Galleries</a></li></ul>
 Exhibitions	<ul style="list-style-type: none"><li>● <a href="#">Exhibitions</a></li></ul>

The manuals can also be exported as PDFs and be saved on the DiSSCo Knowledge Base.



## 5. Recommendations

Basic recommendations can be extracted from the exercise realised to develop the Specialisation tool that lead to a series of actions for 1) further enhancement of the tool and widening of the use of the tool; and 2) its embedment in a wider scenario, under CETAF and DiSSCo umbrellas.

### 5.1 Future use of the tool

The future use of the tool could be the integration of the Specialisation tool in the recurrent activities of the CETAF/ DiSSCO communities with a regular update of the information related to the different categories.

- The technical maintenance of the platform can be achieved by RBINS using the collections servers (Plone, Elastic Search) and a specific Google Drive
- The RBINS/CETAF team needs to continue to develop the technical aspects of the tool with the help of Franck Theeten (Africa Museum). It was proposed as a joint e-service of the Belgian federal DiSSCo/CETAF.
- The DiSSCo RI will define priorities and may guide further modifications of the tool in accordance with new standards such as the Latimer Core for the collection description or the new needs for alignment with the other technical components of DiSSCo.

Role	Definition of the role
Specialisation tool administrator	To be defined by the coordination team
National Node Coordinator	Representative of each country in the DiSSCo RI
Institution manager	To be defined by each institution (optional)
Data provider	Information provider at institution or individual level
Data user(s)	Information search and analysis user at the Individual, Institution NN or DiSSCo RI and external stakeholders with different access levels

Table 5: Roles and definitions

### 5.2 Continued data compilation

The data collection is now the bottleneck of the current exercise.

Many of the information requested in the Google forms are required for the setup of new tools and services. The new site of CETAF prepared with WordPress is totally dependent on the data filled in the Specialisation tool.

The goal was to centralise the collection of data in order to avoid multiple encoding but it is now important to fill the information.

Here is the proposal for the collaborative encoding of the forms by each institution:

<b>Form categories</b>	<b>Recommended profile to fill the information</b>
Collection(s) list	Head of collections
Collection(s) Full Information	Curators, conservators, collection managers
Research Fields	Head of science
Facilities	Coordinator of the facilities
Scientific Instruments	Manager of the scientific instrument
Institution Expertises	Head of Science
Individual Expertise	Scientists, technical experts
Trainings	Member of staff responsible for each training
Galleries	Coordinator of the Galleries
Exhibitions	Coordinator of the Exhibition(s) development

Table 6: Proposed responsible position within the institutions to fill in the information in the different Google forms

In this respect,

- The CETAF Secretariat will coordinate and stimulate the filling and the update of the data by the CETAF/DiSSCo community and help NN to aggregate data at the national levels.

## 6.References

- [ICEDIG deliverable D2.3 Design of a Collection Digitisation Dashboard Deliverable D2.3](#)
- [Synthesys + D2.2 JOINT DASHBOARD OF COLLECTIONS ASSESSMENT TOOLS](#)
- [CETAF Passport and associated collection registry](#)
- [Global Registry of Scientific Collections](#)
- [Latimer Core: A new data standard for collection descriptions](#)

## 7. Appendix : User Manual

The User Manual of the Specialisation plan is composed of 10 guidelines each corresponding to a category/topic and specifically dedicated to the way the user can enter the required data.

Below is the list of the links directing to the manuals of the eponyme topics:

- [Collection\(s\) list](#)
  - [Collection\(s\) Full Information](#)
  - [Research Fields](#)
  - [Facilities](#)
  - [Scientific Instruments](#)
  - [Institution Expertises](#)
  - [Individual Expertise](#)
  - [Trainings](#)
  - [Galleries](#)
  - [Exhibitions](#)
-