DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at info@dissco.eu.

Title
DiSSCo Prepare Deliverable D1.2 Report on Earth sciences use cases and user stories

Author(s)	Identifier of the author(s)
von Mering, Sabine	https://orcid.org/0000-0003-2982-7792
Petersen, Mareike	https://orcid.org/0000-0001-8666-1931
Fitzgerald, Heli	https://orcid.org/0000-0002-6754-6409
Juslén, Aino	https://orcid.org/0000-0001-9434-5250
Raes, Niels	https://orcid.org/0000-0002-4329-4892
Islam, Sharif	https://orcid.org/0000-0001-8050-0299
Berger, Frederik	https://orcid.org/0000-0001-8400-3337
von Bonsdorff, Tea	https://orcid.org/0000-0001-6066-5256
Figueira, Rui	https://orcid.org/0000-0002-8351-4028
Haston, Elspeth	https://orcid.org/0000-0001-9144-2848
Häffner, Eva	https://orcid.org/0000-0001-6448-5826
Livermore, Laurence	https://orcid.org/0000-0002-7341-1842
Runnel, Veljo	https://orcid.org/0000-0001-5198-8678
De Smedt, Sofie	https://orcid.org/0000-0001-7690-0468
Vincent, Sarah	https://orcid.org/0000-0002-4012-0571
Weiland, Claus	https://orcid.org/0000-0003-0351-6523

Affiliation
Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

Contributors
Schulman, Leif
Järvi, Jani
Frank, Jiři
Loo, Tina
Mergen, Patricia
Semal, Patrick

Publisher

Identifier of the publisher

Resource ID
https://doi.org/10.34960/n3dk-ds60

Publication year
2021

Related identifiers
https://doi.org/10.7479/17qp-ge55 (Data publication)
https://doi.org/10.34960/xhxw-cb79 (D1.1)

Relation type
Another reference document

Is it the first time you submit this outcome?
No

Creation date
01/04/2021

Modification date
22/04/2021

Summary of modification

Version

Minor changes to the text of Appendix 3. 1.1

Citation

von Mering, S., Petersen, M., Fitzgerald, H., Juslén, A., Raes, N., Islam, S., Berger, F, von Bonsdorff, T., Figueira, R., Haston, E., Häffner, E., Livermore, L., Runnel, V., De Smedt, S., Vincent, S., Weiland, C. (2021). DiSSCo Prepare Deliverable D1.1 Report on Life sciences use cases and user stories. https://doi.org/10.34960/n3dk-ds60

Abstract

This Deliverable D1.2 from the project DiSSCo Prepare reports the results of Task 1.2 "Analyse Earth sciences use cases and user stories". A total of 128 Earth sciences user stories and use cases was analysed with a special emphasis on the functional demands and required services for the DiSSCo Research Infrastructure. Use cases were gathered from surveys, publications and personal interviews, they were assigned to one out of seven stake-holder groups. For each use case up to five functional demand categories were assigned. Use case analyses revealed that the most important demands for Earth science collections were 'Metadata on collection or record level', 'Advanced search functionality', 'Data integration' and 'Tools for reporting & statistics'. The socio-economic importance of the use cases is discussed and recommendations for the related ongoing Work Packages and the DiSSCo services development in general are given in this report.

Content keywords
scientific

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP1

Project output
Deliverable

Deliverable/milestone number
D1.2

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
sabine.vonmering@mfn.berlin

H2020-INFRADEV-2019-2
Grant Agreement No 871043

DiSSCo Prepare WP 1 Task 1.2 – Deliverable D1.2 Report on Earth sciences use cases and user stories

WP Lead: Henrik Enghoff (UCPH)

Task Lead: Mareike Petersen (MfN)

Authors: Sabine von Mering (MfN), Mareike Petersen (MfN), Heli Fitzgerald (Luomus), Aino Juslén (Luomus), Frederik Berger (MfN), Tea von Bonsdorff-Salminen (Luomus), Niels Raes (Naturalis), Sharif Islam (Naturalis), Rui Figueira (U Lisboa), Elspeth Haston (RBGE), Eva Häffner (BGBM), Laurence Livermore (NHM), Veljo Runnel (UT), Sofie De Smedt (Meise BG), Sarah Vincent (NHM), Claus Weiland (SGN)

Contributors: Leif Schulman (Luomus), Jani Järvi (Luomus), Jiří Frank (NM), Tina Loo (Naturalis), Patricia Mergen (Meise BG), Patrick Semal (RBINS)

Abstract

This Deliverable D1.2 from the project DiSSCo Prepare reports the results of Task 1.2 "Analyse Earth sciences use cases and user stories". A total of 128 Earth sciences user stories and use cases was analysed with a special emphasis on the functional demands and required services for the DiSSCo Research Infrastructure. Use cases were gathered from surveys, publications and personal interviews, they were assigned to one out of seven stakeholder groups. For each use case up to five functional demand categories were assigned. Use case analyses revealed that the most important demands for Earth science collections were 'Metadata on collection or record level', 'Advanced search functionality', 'Data integration' and 'Tools for reporting & statistics'. The socio-economic importance of the use cases is discussed and recommendations for the related ongoing Work Packages and the DiSSCo services development in general are given in this report.

Key words

Collections, DiSSCo, Earth sciences, geosciences, geology, mineralogy, paleontology, gap analysis, cluster analysis, research, use cases, user requirements, user stories, functional demands, service development framework, societal challenges H2020, socio-economic impact/indicators

H2020-INFRADEV-2019-2
Grant Agreement No 871043

INDEX

01 INTRODUCTION

The Distributed System for Scientific Collections (DiSSCo) aims to provide a one-stop-shop for Natural Science Collections objects and associated information in Europe. The planned Research Infrastructure (RI) will be an important source of information for scientists from natural science disciplines but also other users from the sectors education, culture, society, politics, and economy. In order to meet the requirements of all potential stakeholders, the planning and construction of the DiSSCo RI is strongly user driven, especially in the project DiSSCo Prepare.

DiSSCo Prepare Work Package 1, Tasks 1.1 and 1.2 examined the needs of different stakeholder groups for the information that natural science specimens and collections contain and the requirements these needs set for the services to be provided by DiSSCo. More closely, within Tasks 1.1 "Analyse Life sciences use cases and user stories" and 1.2 "Analyse Earth sciences use cases and user stories", we built on existing studies and compilations covering DiSSCo-related use cases and user stories. Task 1.1 and Task 1.2 were complementary to each other, focussing on the two domains Life sciences and Earth sciences, respectively. While Task 1.1 dealt with biological collections (entomological, other zoological, botanical and mycological collections), the focus of Task 1.2 was on collections of fossils, rocks, sediment structures, minerals, and extra-terrestrial material (meteorites).

This report is part of the Deliverables D1.1 "Report on Life sciences use cases and user stories, with recommendations to WP5 and WP6" and D1.2 "Report on Earth sciences use cases and user stories, with recommendations to WP5 and WP6". The complemented corpus of Life sciences and Earth sciences user stories and use cases was analysed with a special emphasis on the functional demands for DiSSCo and its services, as well as their socio-economic importance.

02 APPROACH

As a first step, existing user stories and use case compilations and other resources were collected in a project-wide collaborative effort. All DiSSCo Prepare WP and task leaders as well as the partner institutions working on Tasks 1.1 and 1.2 were contacted and asked to add surveys, presentations, and other sources for user stories to a shared document. An overview of the resources gathered is given in Appendix 1.

From this large compilation, a table of user stories and use cases was generated. It was built mainly on results from the ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage, https://icedig.eu/) effort (van Egmond et al. 2019), but other sources of use cases were added as well (see Appendix 1 for details). To collect and present the use cases, we decided to use the epic story format used e.g. in requirement management and adopted by van Egmond et al. (2019) as well. This format contains the following four parts:

"**As a** [position]… **I want to**… **So that I can**… **For this I need**...".

The compiled table was then adjusted to fit the task's focus on Life sciences or Earth sciences. Accordingly, a number of strictly Life science or Earth science related use cases were removed and others adapted to fit the respective domain focus. Next, duplicates were removed (user stories from different sources that were the same) and near-duplicates (e.g., differing by only one of the stages of the epic story format) were fused without losing information. Incomplete user stories, with no text in 'So that I can...' or 'For this I need...' parts, were also removed.

Subsequently, the uses cases were grouped into the seven user groups or use categories, which were also adopted from the ICEDIG project (van Egmond et al. 2019):

1. **Research** (academic, non-academic incl. Citizen Science)
2. **Collection management**
3. **Technical support** (IT & IM)
4. **Policy** (institutional, national & international)
5. **Education** (academic & non-academic)
6. **Industry**
7. **External** (media & empowerment initiatives)

The collected information was then evaluated, especially regarding existing gaps related to certain stakeholder groups. Once gaps were detected, information on user communities and stakeholders that might help to fill the gaps were collected.

This approach was supplemented by over 15 years of data derived from the SYNTHESYS Transnational Access programme. The programme's record of facility and collection usage requirements and formally published research outputs was clean, refactored and analysed to aid quantification and prioritisation of the user stories previously described. The detailed approach is described in Appendix 3 and summarised in the main discussion and outlook section.

Targeted Groups for additional surveys and interviews

To extend the compilation of use cases and to fill existing gaps, we reached out to all task partners to identify potential users/user groups and stakeholders that should be approached. In a first step, public relations and marketing teams of a number of partner institutions were contacted. Some additional use cases were included from responses by scientists, colleagues and other stakeholder groups.

Earth scientists (geoscientists) and broader stakeholders within the natural science collections community were contacted via mailing lists, directly (via email), targeted surveys, and interviews. In addition, a number of scientific associations/societies and interest groups were contacted. We also contacted federal or other government institutions such as geological services. Appendix 4 provides an overview of targeted groups that were contacted and asked for additional use cases.

Functional demands

During a joint session on "Use cases and user stories" during the virtual All Hands Meeting of DiSSCo Prepare in January 2021, members of both task groups worked together to further analyse the collection of use cases and user stories. The working session focussed on the functional demands these use cases will put on the DiSSCo research infrastructure (RI) and its services.

Some user requirements were already known from the epic user story format part "For this I need". However, the information given there is in many cases rather general and unspecific and functional demands for the DiSSCo Research Infrastructure had to be specified. All task partners worked collaboratively on categorization and harmonization of functional demands resulting from the use cases. Categories and subcategories for functional demands were listed and short definitions explaining what a category or subcategory comprises were included. Up to five functional demand categories or subcategories were allocated for each use case.

Further steps were also discussed during a consultation with project partners from WP5 “Common Resources and Standards” and WP6 “Technical Architecture & Services provision”, and with the DiSSCo Technical Team.

Analysing the user stories

Up to five functional demands were identified for each of the 443 Life science (LS) and Earth science (ES) user stories and converted into a presence/absence matrix. The LS and ES user stories were analysed separately, some user stories applied to both domains. The ES demand categories that were not represented were excluded from the analysis, resulting in a matrix of 128 x 29 (rows & columns). For each category it was scored how often it was scored for the use cases/user stories.

The dissimilarity matrix was calculated for the LS and ES presence/absence matrices separately using the function vegdist from the R-package ‘Vegan’ (Oksanen et al. 2020). As a measure of dissimilarity the Jaccard index was selected and data were subjected to presence/absence standardization. The dissimilarity matrices were further analysed with a hierarchical cluster analysis using an UPGMA (average) cluster algorithm and visualized using the ‘Dendextend’ R-package (Galili 2015). A heatmap visualisation was conducted to assess which similarities in demands result in clusters of user stories. The R-script used in the analysis of the user stories is available in the GitHub repository for the user stories (Raes 2021).

Making the user stories available for future use

All use cases and user stories were imported to GitHub using a semi-automatic import routine. In the repository, the use cases are available as separate “issues” and linked also to the use categories, i.e. the user or stakeholder groups. The functional demand categories scored for each use case were added as separate tags, thus allowing easy filtering.

In addition, the tables comprising all use cases (Life sciences and Earth sciences together as well as separately) were also made available as csv files in a data publication (Fitzgerald et al. 2021). In these tables, numbers were added as simple identifiers (IDs) to identify the different use cases and user stories for later re-use and reference (Appendix 2; Fitzgerald et al. 2021).

03 RESULTS

Compilation of use cases and user stories

The selected, adapted, and sorted user stories and use cases were compiled in a table containing all use cases plus separate tables for Life and Earth sciences (Fitzgerald et al. 2021). The separate tables for both domains are also attached as a supplement to this Deliverable report (Appendix 2).

For Life sciences, a total of 597 user stories were collected (Fitzgerald et al. 2021). The categories "Research" and "Collection management" were the categories with the highest numbers recorded (271 and 173 user stories, respectively). Of the 597 user stories, 33 were gathered from the literature (Vissers et al. 2017, Borsch et al. 2020) and from the Final report of the GBIF Task Group (Krishtalka et al. 2019). After deduplication and removal of incomplete user stories, the number of use cases was 317 (Fitzgerald et al. 2021).

A total of 122 user stories applicable to Earth sciences were collected. The total number of user stories and use cases for Earth sciences after de-duplication and addition was 128 (see Appendix 2; Fitzgerald et al. 2021). With 38 and 48 user stories respectively, the categories "Research" and "Collection management" were also the categories with the highest numbers recorded.

While the high number of user stories in the "Research" and "Collection management" categories was not unexpected, the number of use cases collected in some of the other categories were fewer than expected. The largest potential gaps were detected in the categories "External (media & empowerment initiatives)", "Industry" and Technical support". For the categories "Education", "Technical support" and "Policy", smaller numbers of use cases were also collected. Figures 1 and 2 summarize the number of use cases collected per use case category for both Life sciences and Earth sciences.

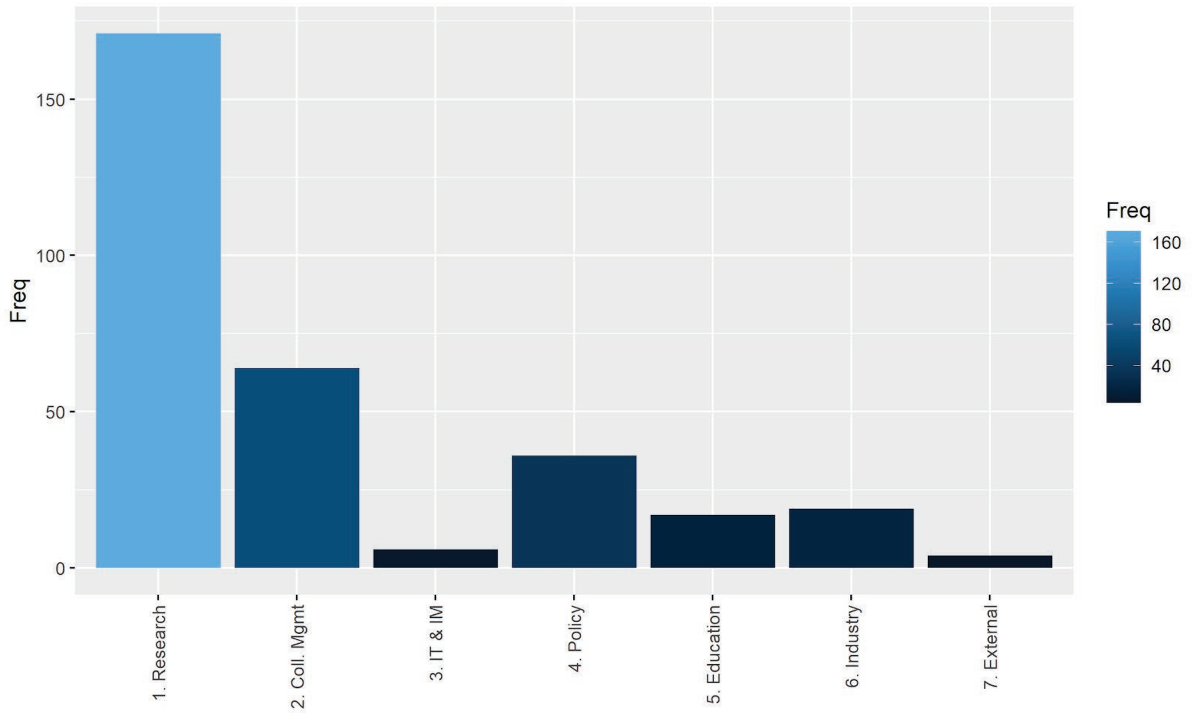

Figure 1. Number of use cases/user stories collected per category for Life sciences, total number of use cases n= 317.

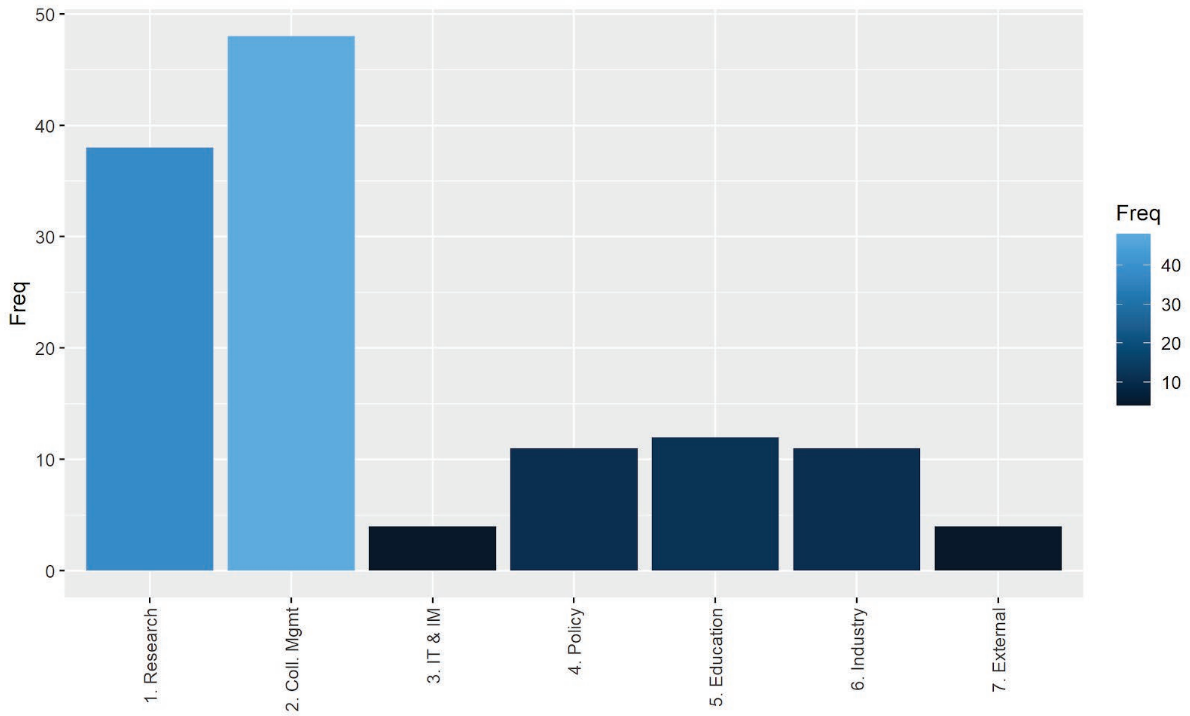

Figure 2. Number of use cases/user stories collected per category for Earth science, total number of use cases n= 128.

Functional demands for the user stories

Communication with the representatives of the DiSSCo Technical team developing technical aspects and services of RI DiSSCo played an important role in ensuring the usefulness of the functional demand categories and recommendations. The development led to recognition of categories and subcategories. Many use cases included more than one functional demand, which were separated.

Altogether 35 functional demand categories and subcategories were identified based on the user stories (Table 1). Functional demands provided by users were placed into 11 categories and 24 subcategories linked to them. The most diverse categories were data and tools with six and twelve subcategories. Definitions for all categories and subcategories are given in Table 1. In the Earth sciences use cases subset, only 29 of the 35 functional demand categories and subcategories were used.

We recognized that data security and API were general demands but not pointed out in the use cases. Therefore they were not scored, except a specific 'Data security' use case requiring limited data access/availability. For life sciences the most important functional demands were 'Tools for data discovery' (scored for 100 of the 317 total use cases), 'Distribution data' (75), 'Morphological data' (59) and both metadata subcategories (58, 54) (Figure 1). For Earth sciences, the five most important functional demand categories were 'Metadata on collection level' (scored for 42 of the 128 use cases), 'Advanced search functionality' (31), 'Data integration' (30), 'Tools for reporting & statistics' (29), 'Metadata on record level' (23) (Figure 2). There were also several functional demand (sub-)categories that were mentioned only once (see Figure 3. and 4.). Even though they were less often required, some of these use cases and the corresponding required functionalities might be relevant for future developments.

Table 1. Functional demand categories (in bold text) and subcategories (normal text) with definitions

Functional demands categories and subcategories	**Definitions**
Advanced search functionality	Advanced search functionalities include technologies like faceted search (filtering), elastic search and Apache Solr
Data	Units of information relating to a specimen or observation
Biochemical or geochemical data	Data describing the biochemical or geochemical composition of a specimen, including secondary compounds in plants
Distribution data	Data describing the specimen collecting locality or observation data
Ecological data	Data describing the original environment and interactions of a specimen or observation, including habitat, associated species or traits
Isotopic data	Data describing the isotopic signature of specimens resulting from isotope analysis
Molecular data	Data describing the molecular composition of a specimen, including DNA/RNA sequence data
Morphological data	Data describing qualitative or quantitative morphological characteristics of a specimen or observation, including measurements
Data integration	Linking of data from different sources, incl. cross-domain (interoperability is a prerequisite); e.g. linking type specimens with the protologue and publications with the specimens used in analyses
Data security	System properties which protect from illegal and malicious data use or from intentional corruption of data systems
Images	Digital representation of specimen images
2D images	Storing and retrieving two-dimensional digital representation of specimens
3D images	Storing and retrieving three-dimensional digital representation of specimens (3D models)

Images related to collections	Storing and retrieving digital images of field notebooks, catalogues, correspondence, photos of sampling area etc.
Label images	Storing and retrieving digital images of specimen labels
Interoperability	Standards and functionality securing interoperability with external services e.g. GBIF, CoL, thesauri
Legal and policy framework	Rules and procedures related to legal and policy issues, such as access policies or information on legal obligations linked to specimens, but also standardized information on the use assets within the infrastructure
Metadata	Information describing or providing additional facts for any part of the data
Metadata on collection level	Information describing or providing additional facts for a set of specimens
Metadata on record level	Information describing or providing additional facts for a single record
Physical access	Physical access to collections e.g. sub-collections, certain specimens
Reference system & Standard lists	Description of underlying data standards, data architectures and vocabularies which are needed for collection information system management and integration with other systems (e.g. GeoNames, Global Names Architecture, GBIF Backbone Taxonomy)
Tools	Digital applications to perform various tasks with specimen data or to interact with databases
Annotation tools	Tools to add additional information to specimen data
Tools for clustering requests	Tools to cluster related requests based on information including requester, purpose and material requested in communication/feedback system
Tools for data analysis	Tools to create data quality assessments, species distribution models
Tools for data discovery	Tools to perform complex specimen data searches simultaneously from different collections
Tools for data visualization	Tools to visualize data
Tools for documentation	Tools to enable documentation of collections and collection history

Tools for downloading data/metadata	Tools to download data or metadata resulting from a data discovery event
Tools for geo-referencing	Tools to perform georeferencing of localities relating to collections
Tools for identification	Tools to view or capture trait data for identification or tools for automated identification of specimens
Tools for limiting access to data	Tools to restrict access to any part of the data relating to a specimen, due to cultural or environmental sensitivity
Tools for reporting & statistics	Tools to produce structured data in the form of reports or statistical summaries
Tools for uploading	Tools to upload or import data

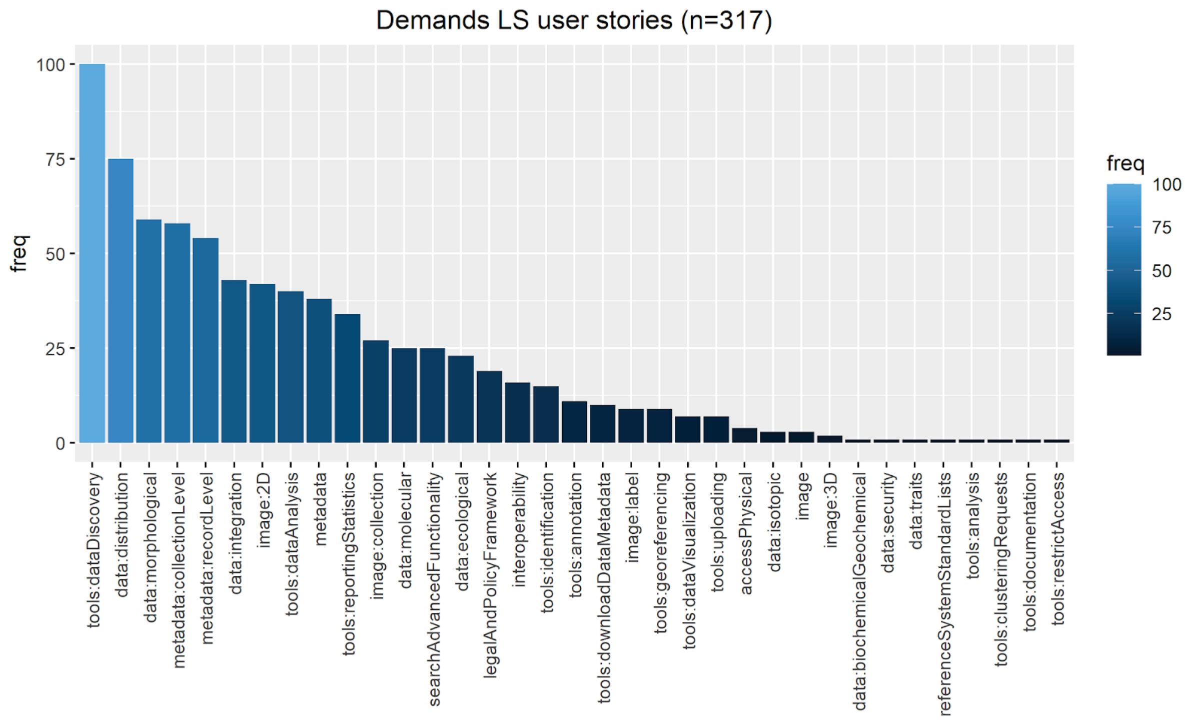

Figure 3. Number of times each of the 35 categories of functional demands was mentioned in Life Science (LS) user stories.

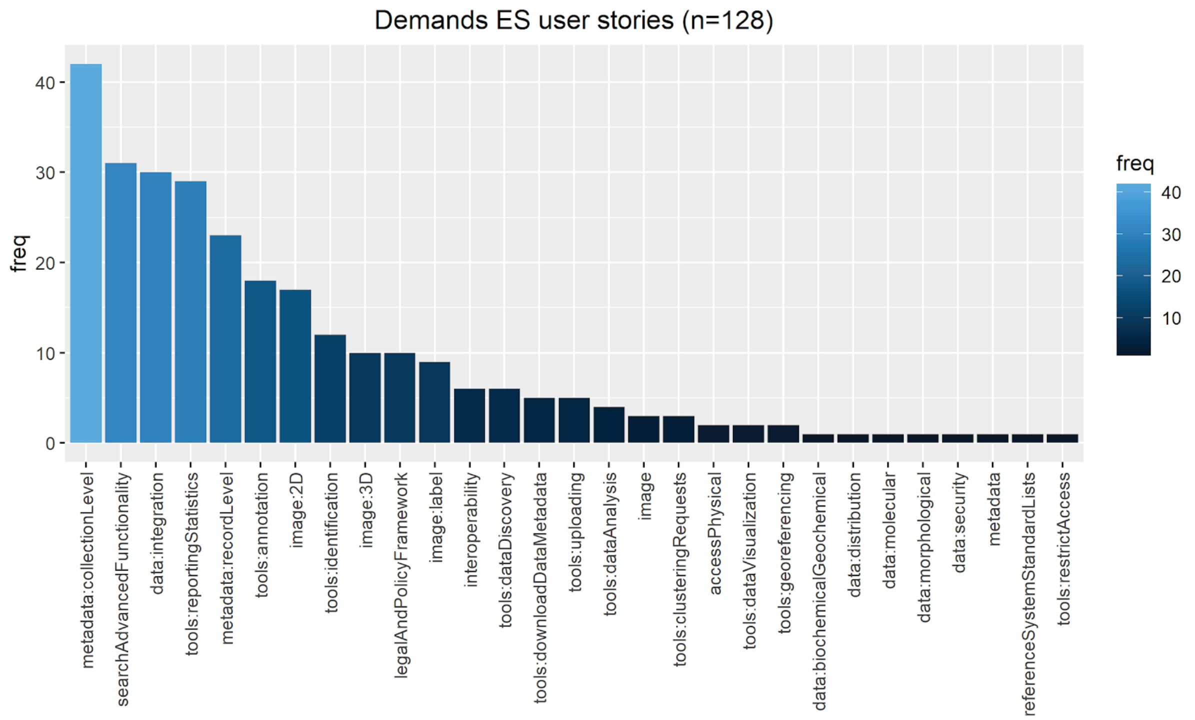

Figure 4. Number of times that each of the 29 functional demand (sub-)categories was mentioned in Earth sciences (ES) user stories.

Analysing the user stories

The cluster analysis of the Earth science use cases/user stories resulted in nine different cluster groups, shown in different colours in Figure 5 (also available in high resolution in Fitzgerald et al., 2021, which allowing to zoom in). These cluster groups correspond to use cases grouped together by having the same or a combination of similar functional demand categories (see heatmap in Figure 7). As an example, we looked at the blue cluster, which is marked by a red rectangle in Figure 5 and enlarged in Figure 6. The clustering functional demands in the two sub-cluster of the blue cluster are 'Metadata on collection level' or the combination of 'Metadata on collection and record level', and a combination of 'Metadata at collection level' together with 'Tools for reporting & statistics'.

The full list of functional demands behind the user story IDs and cluster dendrogram can be found in Appendix 2 and in Fitzgerald et al. (2021).

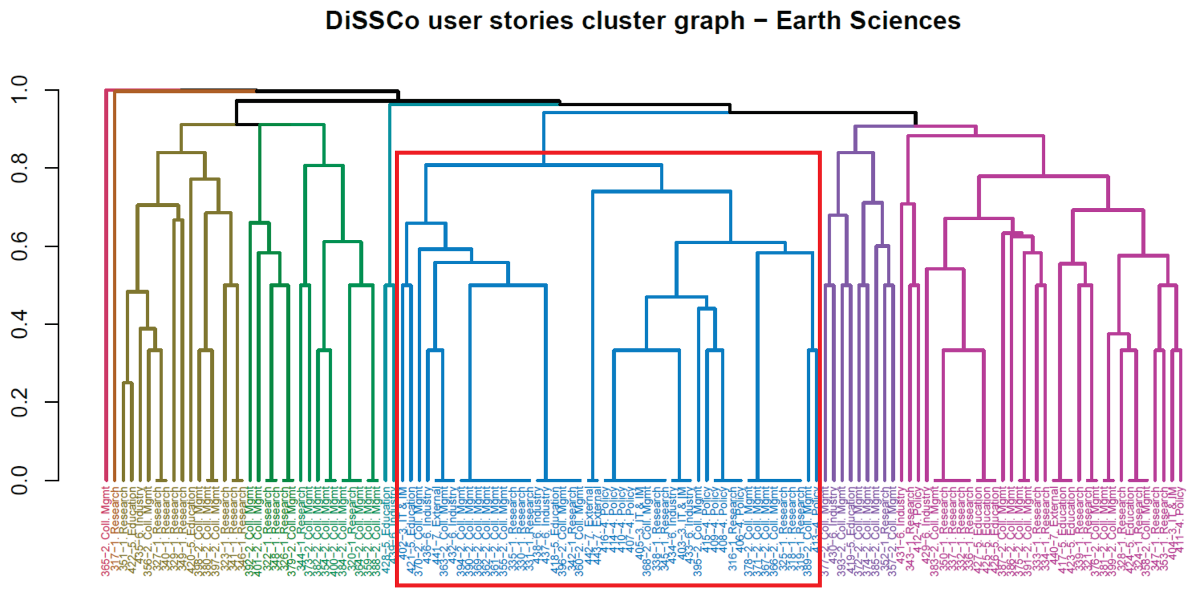

Figure 5. Cluster dendrogram for the Earth Sciences user stories (n=128). Different colours correspond to nine different cluster groups. Use case IDs and user categories are given.

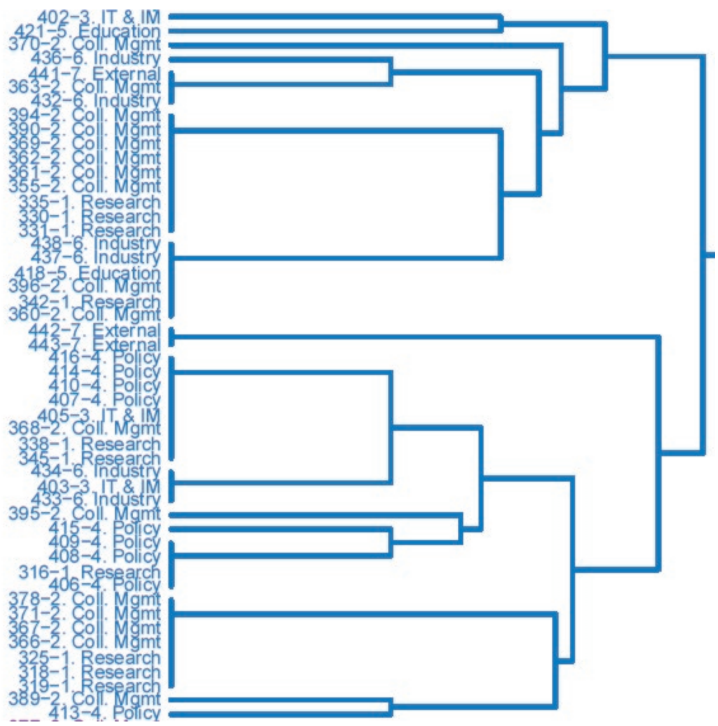

Figure 6. Cluster group example (area shown as red rectangle in Figure 5), use case IDs and user category are given.

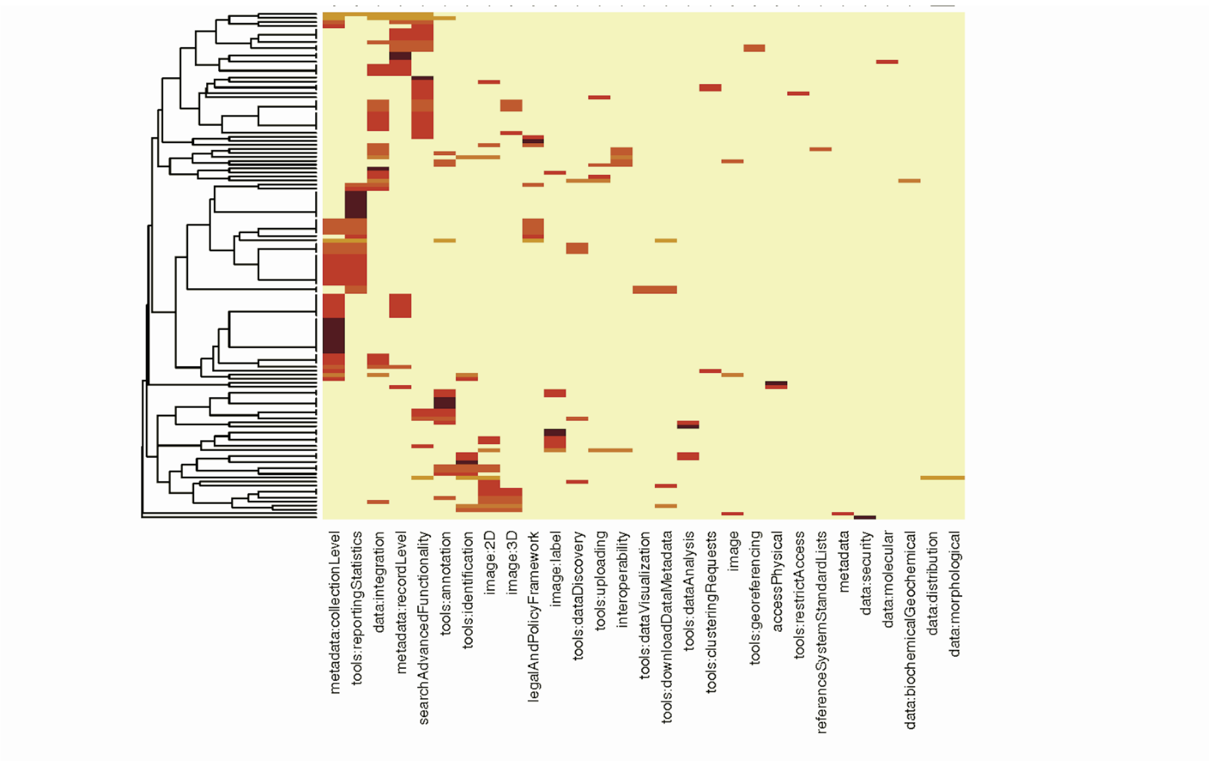

Figure 7. Heatmap Earth Sciences user stories (n=128). Left cluster dendrogram corresponds to Fig. 6. The x-axis shows the 29 reported functional demands.

Making the user stories available for future use

To allow easy reuse, the user stories and use cases including the functional demands were made available in different formats. A data publication comprises the tables with the use case IDs, user group/use categories, descriptions of the use cases in the 'epic format', functional demand (sub-)categories plus the figures of the use case analysis (Fitzgerald et al. 2021).

In addition, all use cases and user stories incl. functional demands were made available in a dedicated repository on the platform GitHub (https://github.com/DiSSCo/user-stories) as a "collection of user stories describing evolving requirements of stakeholders involved in managing and using natural science collections". This facilitates future reuse of the whole compilation or selected use cases and allows referring to them separately or as a collection e.g. during the development of specific tools. Figure 8 shows two example use cases as presented in the GitHub repository.

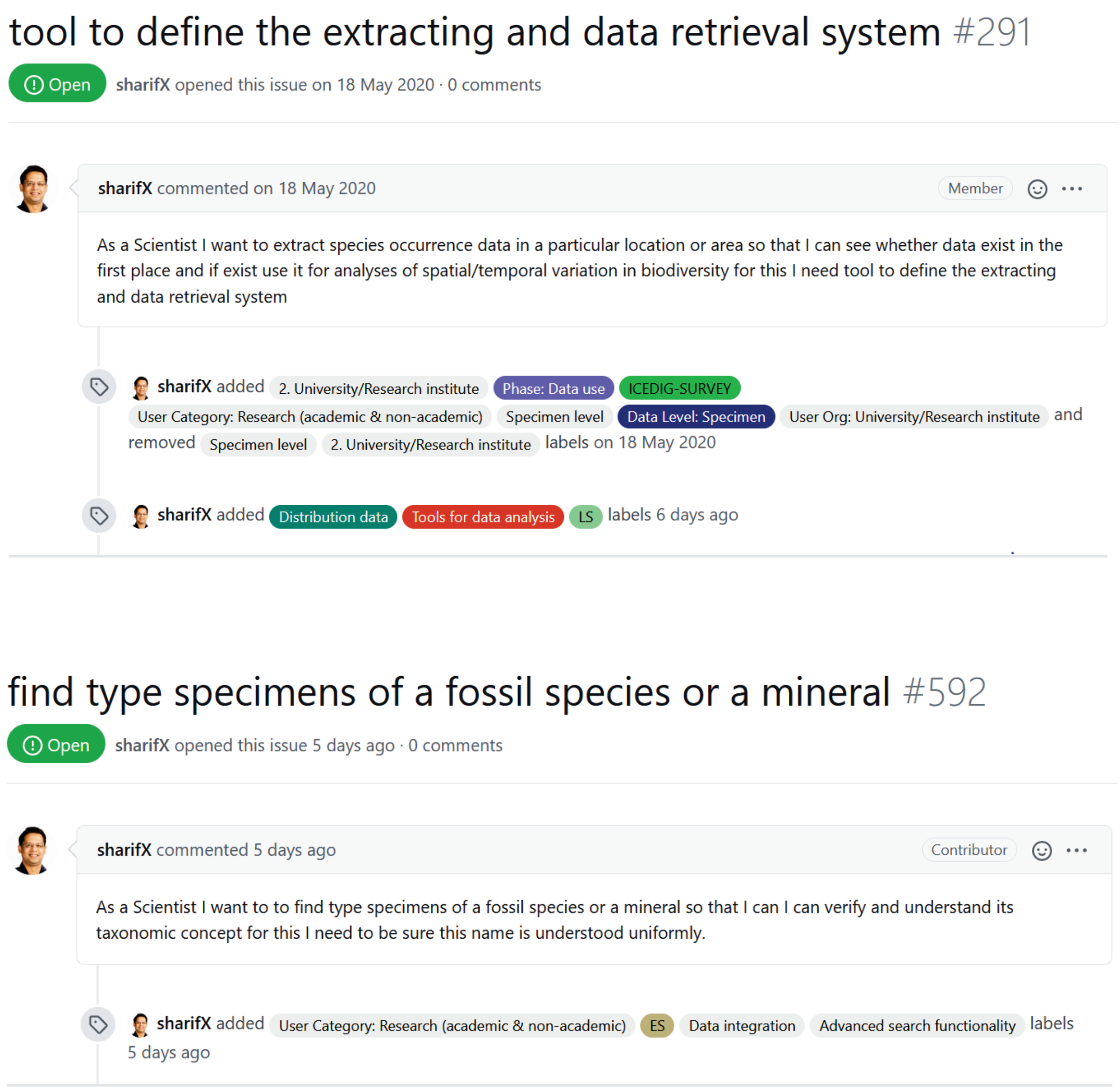

Figure 8. Screenshot of two selected use case in GitHub repository, shows tags for functional demands / use categories etc.

04 DISCUSSION AND OUTLOOK

Use cases and functional demands

The categories of use "Research, Collection Management, Technical support, Policy, Education, Industry, and External" were represented, the far most use cases representing Research and Collection Management.

The societally wide-ranging needs for the use of scientific collections came to the fore (Figure 1.). For instance, high-quality metadata descriptions and images are highly needed (Figure 3 and 4) and serve stakeholders from research to industry (Fitzgerald et al. 2021). The recognition and description of functional demands at appropriate and useful levels required several rounds of refinement to optimise usefulness for the further development within DiSSCo (see Table 1). The analysis of use cases and recognition of functional demands create basis for and support further RI DiSSCo development, e.g. recognition of digitisation prioritisation criteria (Task 1.3) and set the service development framework.

Data derived from records of recent collection-based projects and their outcomes can be used to augment and quantify the user stories gathered. The majority of projects submitted under existing (in-person) collection access schemes require the use of both collection material and local analytical facilities such as molecular labs, microanalysis and 2D/3D imaging suites, and geochemical identification (see Appendix 3 for details). For facilities/services that can be delivered remotely, this information is key to developing robust, appropriately scaled digital services and workflows.

Use cases and societal challenges in Europe 2020 strategy

The analysis of user stories, including its functional demands, provides important information for the identification of socio-economic impact of DiSSCo. The framework for this analysis (under development in DiSSCo Prepare Task 1.4) will link the impact of uses and applications of DiSSCo and related activities to a set of major objectives, defined by the ESFRI Working Group on monitoring RI performance (ESFRI 2019). These include:

- Enabling scientific excellence
- Delivery of education and training
- Enhancing transnational collaboration in Europe
- Facilitating economic activity
- Outreach to the public
- Optimising data use
- Provision of scientific advice
- Facilitating international co-operation
- Optimising management

These can also be arranged within four major areas - scientific excellence, economy and innovation, society, and policy - contributing to the following H2020 societal challenges (Societal Challenges, Horizon 2020):

- Health, demographic change and wellbeing;
- Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the bioeconomy;
- Climate action, environment, resource efficiency and raw materials;
- Europe in a changing world - inclusive, innovative and reflective societies;

We identified user stories which directly or indirectly address some of the societal challenges which are outlined in Horizon 2020 strategy documents. We also took into account the user stories from (van Egmond et al. 2019) earlier DiSSCO findings from the ICEDIG project. The user stories are summarized as follows.

Health, demographic change and wellbeing

Occurrence data of virus/bacterial disease vector species (mosquitoes, bats, etc.) can be obtained from collection databases to support control of zoonotic diseases. Comprehensive, connected collection management systems can make remote working easier for researchers, minimizing the risks associated with pandemics. Biodiversity data from collections can be linked to health data so that health impact of variables such as vegetation cover, deforestation or species diversity can be modelled. Future distributions of allergenic species can be modelled by linking collection and occurrence data with e.g., climate data/environmental data, helping to predict and mitigate health problems arising from allergies. Collection-based research can support ethnobotanical research, as seen in an investigation of natal diets (de Boer & Lamxay 2009).

Food security

Genomic data from collections can be used to analyse food plant diseases, see Essakhi et al's study of Phaeoacremonium species *(hyphomycetes fungi) and their association with esca disease in cultivated grapevines* (e.g. Essakhi et al. 2008). Specimens can support the identification, conservation and use of crop wild relatives. They can also provide occurrence and associated environmental data which can contribute in investigating genetic control of agriculturally relevant traits in crop wild relatives.
Locating the origin of food pest species and modelling their future distribution by e.g. linking occurrence data with climate data. Support crop diversification and green infrastructures to increase sustainability of farming practices, investigate genetic control of agriculturally relevant traits in crop wild relatives.

Inclusive, innovative and reflective societies

Digital species information such as images, trait descriptions, etc., can help citizen scientists to identify species and facilitate citizen science projects. Digitally opening up natural history collections for education and public knowledge can facilitate engagement with underrepresented citizen groups in citizen science. Runnel et al. (2019) studied the role of natural history collections in improving the digital skills of citizens.
In natural history museum educational programs, the use of digital content (including digitized specimens) was considered underdeveloped with a high potential for increased use and effectiveness (Runnel et al. 2019). They propose that enhancing data search and building public interfaces for collection digital content could lead toward more effective use of natural history digital content in society and broader acknowledgment of the value of natural history collections. Including citizen scientists in museum-based citizen science projects (particularly digitisation tasks) can also lead to involvement of participants in decision-making regarding environmental topics in society (Runnel et al. 2019). As implied by Runnel & Wijers (2019), citizen science attribution in crowdsourcing project outcomes needs to be improved, often the published data lacks the information about citizens involvement .
Private natural history collections are historically and practically an important part of biological research. In some cases, private collections are donated to museums or other collection holding institutions, but often their scientific value is lost due to loss or deterioration of the collections. If private collection holders are invited to share the specimen data, the impact to science will be achieved quicker and also the handling of collections in case of donations will me much smoother. Opening collection data management systems to private collectors can also put greater emphasis on the value of citizen science and its part in academic research. Willemse et al. (2020) analysed the perspective of private collectors and proposed tools and recommendations for the DISSCO consortium.

Climate action, environment

Well-documented natural history collections can be of great value for climate change research (Johnson et al, 2011). Access to historical data of species distributions, abundance and habitats can help to understand climate change. Metadata (sampling methodology, etc.) is crucial for specimen data useability and applying statistical approach. Easy and comprehensive data access can also facilitate innovative data solutions for visualisation of species loss and environmental degradation and have an impact on policy making. Digital specimen data will allow aggregation of specimen data to other data types, such as climate and weather data and land-use data will allow new types of analysis serving e.g. climate change mitigation at ecosystem level. Digital specimen trait data can be used to study links between species morphology and climate change (Salinas-Ramos et al, 2020).

05 RECOMMENDATIONS AND LINKS TO OTHER WORK PACKAGES

The use cases and functional demands can help prioritize developments in the Technical Work Packages of DiSSCo Prepare. All identified use cases were imported into GitHub, the main repository for technical developments in DiSSCo Prepare and related projects (DiSSCo/user-stories). Therewith, they are easily accessible for the developmental teams and can be taken into consideration when setting up developmental plans for the DiSSCo technical architecture, the prioritization of services and setting up or shaping pilots in DiSSCo Prepare.

The results of WP 1 provide a valuable resource for WP 3, which develops the specifications for a digital maturity self-assessment tool in task 3.1. The analysis of user stories of WP1 may indicate areas where digital maturity of consortium members will be needed to underpin the success of the DiSSCo services. Furthermore, it has to be presumed that any assessment referring to actual user needs potentially experiences a higher acceptance during implementation. In a first attempt to identify areas that provide information about digital maturity a subset of 127 user stories from WP1.2 were categorized and described in a similar approach as the one adopted to identify functional demands. As a first result, a non-exhaustive list of areas to be covered was derived from the dataset. According to this interpretation of user needs the following areas are of primary interest and should therefore be considered, when assessing digital maturity of institutions or infrastructure.

- Availability of data at collection and specimen level
- Data standards and quality assurance
- Licensing
- Open data policies and processes (links to WP2)
- Availability of data types e.g. 3D
- Progress of digitisation, plus workflows, best practices, etc.
- Infrastructure including collections management systems
- Persistent identifiers
- Analytics and monitoring
- Availability of tools and processes e.g. annotation, transcription, AI, etc.

It is planned to extend the analysis of WP3 to the whole set of user stories and to take the functional demands categories and sub-categories into account. This would lead to a higher degree of standardization. However, it has to be evaluated, if the functional demands define useful categories for the specifications of a digital maturity assessment tool.

For the development of “Common resources and standards” (DiSSCo Prepare WP5) the following recommendations should be taken into account:
DiSSCo Knowledgebase (Task 5.1): Identified functional demands should contribute to the content in the DiSSCo Knowledgebase. Here, e.g. compilation of necessary (meta) data standards on object and collections level should be available, relevant policies should be presented and described (ongoing collaboration with Task 7.3 DiSSCo Policies), and available tools to address user needs should be listed and instructions need to be given. In addition, reporting on collections-based research but also use cases from other stakeholder groups could be documented here.

Not all identified functional demands are directly linked to DiSSCo’s central architecture (Core Digital Object) but rather to services and products (e.g. for data publication, reference systems) linked to it. The starting Task “Technical infrastructure for science data mobilisation and publication” (Task 5.4) need to consider the identified requirements for development plans of e.g. Catalogue of Life and GeoCASe.

The development of “Technical Architecture and Service Provision” (DiSSCo Prepare WP6) should consider the identified functional demands, translate them into a more technical language, and use them for prioritization in the technical development of the architecture. For example the list of required tools for end users (compare Table 1) should be harmonized with current concepts of the DiSSCo RI. However, not all services and tools might be a necessary development under DiSSCo and we recommend WP6 partners to make clear which tools and services are already available or planned in associated RI (Task 6.4 “Embedding Embedding DiSSCo in the technical landscape”).

The use cases and functional demands will directly contribute to Task 6.1 “CMS systems interoperability and harmonisation”. Within this task, which involves harmonization, specifications and agreements for local collection facilities to achieve interoperability with DiSSCo’s emerging core infrastructures, a modeling framework adopting the EventStorming format was created to capture main events in the life cycle of a Digital Extended Specimen (DES). Based on a lightweight common description model - “a command causes an event, which can lead to a reaction” - more formal representations of DES-related processes and activities like initial digitization, assignment of PIDs, further (sub)sampling for genomic analysis or taxonomic revisions were developed.

These representations will now be used in the context of Task 6.1 for the implementation of event data in a common specification like CloudEvents to provide interoperability across DiSSCo-linked services, platforms and systems, but should also be connected to corresponding use cases and user stories. Aim is the establishment of a common notation to gain a unified understanding of needs as well as “responding” technical solutions following the aforementioned “Command -> Event -> Reaction” scheme. The utilization of such a unified and consistent framework for requirements, objectives and services will substantially support the full implementation of the DiSSCo service architecture.

In addition, the DiSSCo system design should anticipate future reporting on impact metrics: consistent, ongoing categorisation of collection usage/data access requests and the implementation of data quality controls on downstream research output metadata such as peer-reviewed publications will facilitate linkages to existing data sources in the wider scholarly publications ecosystem, improving the reliability and usefulness of impact metrics.

Data integrity and interoperability would be improved by incorporating existing, discipline-specific digital services in the development of further digital collection systems. For example, incorporating vocabularies derived from taxonomic name lookup and resolution services such as the GBIF Species API would make reporting on life sciences collection access requirements more efficient, granular and repeatable. It would also facilitate identification of data gaps and feed into transnational digitisation prioritisation and planning.

The use of more standardised data and related system linkages will also support the creation of visual analysis and decision-making tools such as dashboards: if such interfaces are to be intuitive and understandable by stakeholders external to the natural sciences community, collection usage and impact data must be optimised suitable for high-level aggregation and visualisation.

Data from in-person access schemes (like Transnational Access in SYNTHESYS) can be used as a benchmark from which we can assess progress towards increased engagement by under-represented groups. If remote access to equivalent collection data and services overcomes barriers to inclusion, this should be measurable in changes to the demographic profile of the user base. This type of data-driven service design would require demographic data on system end-users to be recorded, where appropriate. More inclusive categories of complex demographic variables such as gender identity should also be incorporated: if these are not in the data, the engagement and scientific impact of these groups cannot be measured, reported on or incorporated into ongoing system design and development activities.

In order to track and better understand our users and how they need/want to use our facilities in the future, we should consider providing solutions for the following: (i) Enabling scientific excellence, (ii) Delivery of education and training, (iii) Enhancing transnational collaboration in Europe, (iv) Facilitating economic activity, (v) Outreach to the public, (vi) Optimising data use, (vii) Provision of scientific advice, (viii) Facilitating international co-operation, and (ix) Optimising management.

Many tools and services are already available and meet at least partly the requirements of known and new users of Natural Science Collections. However, there is still potential for improvement, especially in linking services from different RI and making tools interoperable. The DiSSCo RI need to accept this challenge and provide solutions for use cases identified and described in this deliverable report. Especially those which require more than only one functional demand and those which can only contribute to our societal challenges with a comprehensive set of services linked to our Natural Science Collections.

06 REFERENCES

de Boer, H. & Lamxay, V. (2009). Plants used during pregnancy, childbirth and postpartum healthcare in Lao PDR: A comparative study of the Brou, Saek and Kry ethnic groups. J. Ethnobiology Ethnomedicine **5:** 25. https://doi.org/10.1186/1746-4269-5-25

Borsch, T., Stevens, A.-D., Häffner, E., Güntsch, A., Berendsohn, W.G., Appelhans, M.S., Barilaro, C., Beszteri, B., Blattner, F.R., Bossdorf, O., Dalitz, H., Dressler, S., Duque-Thüs, R., Esser, H.-J., Franzke, A., Goetze, D., Grein, M., Grünert, U., Hellwig, F., Hentschel, J., Hörandl, E., Janßen, T., Jürgens, N., Kadereit, G., Karisch, T., Koch, M.A., Müller, F., Müller, J., Ober, D., Porembski, S., Poschlod, P., Printzen, C., Röser, M., Sack, P., Schlüter, P., Schmidt, M., Schnittler, M., Scholler, M., Schultz, M., Seeber, E., Simmel, J., Stiller, M., Thiv, M., Thüs, H., Tkach, N., Triebel, D., Warnke, U., Weibulat, T., Wesche, K., Yurkov, A. & Zizka, G. (2020). A complete digitization of German herbaria is possible, sensible and should be started now. Research Ideas and Outcomes **6:** e50675. https://doi.org/10.3897/rio.6.e50675

Cech, E. A. & Blair-Loy, M. (2019). The changing career trajectories of new parents in STEM. Proc. Natl Acad. Sci. USA **116:** 4182–4187. https://doi.org/10.1073/pnas.1810862116

DiSSCo user-stories (2021). DiSSCo/user-stories: Collection of user stories describing evolving requirements of stakeholders involved in managing and using natural science collections. GitHub, https://github.com/DiSSCo/user-stories, last accessed 2021-04-01.

van Egmond, E., Willemse, L., Paul, D., Woodburn, M., Casino, A., Gödderz, K., Vermeersch, X., Bloothoofd, J., Wijers, A. & Raes, N. (2019, March 31). Design of a Collection Digitisation Dashboard (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.2621055

ESFRI (2019). Report of the ESFRI Working Group on monitoring RIs performance | www.esfri.eu. https://www.esfri.eu/latest-esfri-news/report-esfri-working-group-monitoring-ris-performance

Essakhi, S., Mugnai, L., Crous, P.W., Groenewald, J.Z. & Surico, G. (2008). Molecular and phenotypic characterisation of novel Phaeoacremonium species isolated from esca diseased grapevines. Persoonia **21:** 119–134. https://doi.org/10.3767/003158508x374385

Fitzgerald, H., von Mering, S., Juslén, A., Petersen, M., Berger, F., von Bonsdorff-Salminen, T., De Smedt, S., Figueira, R., Frank, J., Häffner, E., Haston, E., Islam, S., Järvi, J., Livermore, L., Loo, T., Mergen, P., Raes, N., Runnel, V., Schulman, L., Semal, P. & Vincent, S. (2021). Compilation of use cases and user stories, functional demands and their analyses for the RI DiSSCo – DiSSCo Prepare Tasks 1.1 & 1.2. [Dataset]. Data Publisher: Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. https://doi.org/10.7479/17qp-ge55

Galili, T. (2015). dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. Bioinformatics **31:** 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Johnson, K.G., Brooks, S.J., Fenberg, P.B., Glover, A.G., James, K.E., Lister, A.M., Michel, E., Spencer, M., Todd, J.A., Valsami-Jones, E., Young, J.R., Stewart, J.R. (2011). Climate Change and Biosphere Response: Unlocking the Collections Vault. BioScience 61: 147–153. https://doi.org/10.1525/bio.2011.61.2.10

Krishtalka, L. (2016). Accelerating the discovery of biocollections data. Copenhagen: GBIF Secretariat. Available online at: http://www.gbif.org/resource/83022

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner H. (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Raes, N. (2021). R-script used for analysis of user stories and functional demands. GitHub https://github.com/DiSSCo/user-stories/blob/17315c736931a5c8ce6fb1837f143a7ca6e07658/Raes-N_DPP_T1-1_T1-2.R

Runnel, V., Hardy, H., Sanghera, H., Robinson, L., Shennan, V., Livermore, L. & De Smedt, S. (2019). Natural history collections and digital skills of citizens. Zenodo. https://doi.org/10.5281/zenodo.3364541

Runnel, V. & Wijers, A. (2019). Improving the detection of collection-based citizen science projects. Zenodo. https://doi.org/10.5281/zenodo.3364519

Salinas-Ramos, V.B., Agnelli, P., Bosso, L., Ancillotto, L., Sanchez-Cordero, V., Russo, D. (2020). Body Size Variation in Italian Lesser Horseshoe Bats Rhinolophus hipposideros over 147 Years: Exploring the Effects of Climate Change, Urbanization and Geography. Biology (Basel). **10:** 16. https://doi.org/10.3390/biology10010016

Societal Challenges, Horizon 2020. European Commission. Available online at: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges#Article

Vissers, J., Van den Bosch, F., Bogaerts, A., Cocquyt, C., Degreef, J., Diagre, D., De Haan, M., De Smedt, S., Engledow, H., Ertz, D., Fabri, F., Godefroid, S., Hanquaert, N., Mergen, P., Ronse, A., Sosef, M., Stévart, T., Stoffelen, P., Vanderhoeven, S. & Groom Q. (2017). Scientific user requirements for a herbarium data portal. PhytoKeys **78:** 37–57. https://doi.org/10.3897/phytokeys.78.10936

Willemse, L., Runnel, V., Saarenmaa, H., Casino, A. & Gödderz, K. (2020). Digitisation of private collections. Zenodo. https://doi.org/10.5281/zenodo.3598303

07 APPENDICES

Appendix 1. Overview of existing user surveys, presentations, and other sources collaboratively collected by project partners.

Author(s) (publication year)	Title	URL
van Egmond, E., Willemse, L. & al. (2019)	Design of a Collection Digitisation Dashboard	https://icedig.eu/sites/default/files/deliverable_d2.3_icedig_-_design_of_a_collection_digitisation_dashboard_v1.0.pdf
Raes, N. (2019)	DiSSCo user stories collection	https://dissco.teamwork.com/#/files/8287666
Raes, N. (2019)	DiSSCo user stories presentation	https://dissco.teamwork.com/#/files/8146993
Collection Description Interest Group (2020)	Use Case Analysis	https://docs.google.com/spreadsheets/d/1SsfwogZ88TgouDJ7EoDqXJFol-eVs7aYdFx504qJNzc/edit#gid=0
Collection Description Interest Group (2020)	Use Cases	https://github.com/tdwg/cd/tree/master/reference/use_cases
Anonymous (2020)	User stories for SYNTHESYS Plus T2.2 dashboard	https://docs.google.com/spreadsheets/d/1weWdM_5wCAdr49-rH-8c5fgOrTTb8yYwnHsGpCw_oMo/edit#gid=2125639734
inspired from CETAF Earth Sciences group discussions (2020)	User story for geological specimens	Example of species page: https://www.mindat.org/min-4322.html GeoCase Portal http://www.geocase.eu/access
DiSSCo user stories (2020)	ordered user stories incl. ICEDiG Survey and ELViS Survey user stories	https://github.com/DiSSCo/user-stories/projects/1
TDWG CD user cases	use cases for collection descriptions	https://github.com/tdwg/cd/tree/master/reference/use_cases
Addink W., Belknap G. & al. (2017)	DiSSCo Design Study Report	unclear where published but study is known to the community
Borsch & al. (2020)	A complete digitization of German herbaria is possible, sensible and should be started now	https://riojournal.com/article/50675/
Vissers, J., Bosch, F. van den & al. (2017)	Scientific user requirements for a herbarium data portal	https://phytokeys.pensoft.net/article/10936/
Petersen, M., Hoffmann, J. & Glöckler, F. (2019)	Access to Geosciences – Ways and Means to share and publish collection data	https://riojournal.com/article/32987/
Krishtalka L., Dalcin, E. & al. (2016)	Accelerating the discovery of biocollections data	http://www.gbif.org/resource/83022

Appendix 2. Use cases and functional demands Earth science

ID	CAT	AS (POSITION)	I WANT TO ...	SO THAT I ...	FOR THIS I NEED ...	FUNCTIONAL DEMAND 1	FUNCTIONAL DEMAND 2	FUNCTIONAL DEMAND 3	FUNCTIONAL DEMAND 4	FUNCTIONAL DEMAND 5
316	1	Association	to gather information to have overall figures representative of partners' state-of-the-art	I can showcase the relevancy of the collections held to policy makers and attract funds	high-level figures that feature the collections as a whole	Metadata on collection level	Legal and policy framework	Tools for reporting & statistics		
317	1	Citizen Science (CS) site manager	select multiple images	I can build a CS project	information on the basic elements of the images	Images	Metadata			
318	1	Citizen scientist	be recognized as contributor	I can apply for funding to digitize my own collections	contribution indicators (as contributor)	Tools for reporting & statistics				
319	1	Citizen scientist	be recognized as contributor	I can identify my contribution on validating data from external sources	contribution indicators (as validator)	Tools for reporting & statistics				
320	1	Citizen Scientist	curate and add untranscribed labels	I can contribute to the overall project, perhaps especially on particular group of fossils/rocks etc.	a curation interface	Annotation tools				
321	1	Citizen Scientist	find specific info	I can use it in blogs, publications, etc.	metadata and photo's	Metadata on record level				
322	1	Citizen scientist	help with transcribing	I can enjoy this voluntary work	images without transcription	2D images	Label images			
323	1	Citizen Scientist	identify a geological or palaeontological specimen	I know which rock, mineral, gem or fossil I have encountered	an automated identification tool	Tools for identification				
324	1	Citizen scientist	know where was a certain collector on a certain day	I can help transcribe a specimen	existing transcription of specimens collected around the same time by the same collector	Advanced search functionality	Metadata on record level			

325	1	scientist	be recognized as contributor	I can apply for funding to digitize any collection of special interest for research	contribution indicators (as contributor/validator)	Tools for reporting & statistics				
326	1	Scientist	check that the transcribed label data corresponds to the actual label information	I can confirm collection details	digital images of specimen labels	Label images				
327	1	Scientist	check the identity of the specimen	I can confirm the identification	digital images of rocks/minerals/fossil species	2D images	3D images			
328	1	Scientist	do biographical research on specific collectors	I can contribute to cultural history (e.g. colonial history)	names of collectors and places of origin of specimens	Data integration	Metadata on record level	Advanced search functionality		
329	1	Scientist	download a collection of images with a resolution of 400 dpi in jpeg 2000 format	I can use them as inputs for training a neural network designed to classify similar images	a method to query and download image collections according to a set of parameters	Tools for downloading data/metadata	2D images			
330	1	Scientist	expose the collection I am contributing to other users, according to the FAIR principles	I can collaborate with other users on new science using the collection	an overview of existing collection(s) with standard descriptions	Metadata on collection level				
331	1	Scientist	extract data on presence and storage of all collected specimens of rocks/minerals of my present interest and details of loan process at insitutions that host the material	I can easily choose and loan material for scientific work	database on presence and storage of specimens of rocks/minerals and host institution information on loan process	Metadata on collection level				
332	1	Scientist	find all literature citing a specimen	I can recontruct the history of its classification	links between specimens and literature	Data integration	Advanced search functionality			

333	1	Scientist	find an image of a rock/mineral etc. to put in a press release	I can allow journalists to write an article on my paper	images available with preferably very open CC licence	2D images	Advanced search functionality			
334	1	Scientist	find and reuse digital specimens from DiSSCo	I can use all digital specimens of a selection of rocks/minerals/fossil species from DiSSCo for scientific research	fast access to the DiSSCo infrastructure	Advanced search functionality				
335	1	Scientist	find out if a collection holds specimens of interest	I can study the specimens	collection level data to make estimations about holdings data	Metadata on collection level				
336	1	Scientist	find out if there are additional images available from a particular specimen	I can include different views of a specimen in a written report	a reference to related images for a particular specimen	Data integration	Advanced search functionality			
337	1	Scientist	find out if there is an alternative image available with a resolution of 300 dpi and in png file format	I can select the appropriate image for including in a paper according to the publisher requirements	a reference to the alternative image formats available	Data integration	Advanced search functionality			
338	1	Scientist	find out what is in the collection but not been digitised	I know whether its available for me to use	A high level description of the collection	Metadata on collection level	Tools for reporting & statistics			
339	1	Scientist	find specimens that have DNA data	I can analyze past genetic diversity from subfossil remains	ancient DNA from collection specimens	Molecular data	Metadata on record level			
340	1	Scientist	find whether images available at institutional repositories are of taxonomic-grade	I can use the images as surrogates for specimens, to reduce specimen travel and associated risks	extensive and standardized metadata on the image including photogrammetric data	2D images	3D images			

341	1	Scientist	generate an artificial intelligence (AI) algorithm based on a selection of geological specimens	I can run the algorithm on images of other institutes to verify identifications and recover further specimens	an AI toolbox to generate the AI algorithm	Tools for identification	Tools for data analysis			
342	1	Scientist	know all rocks/minerals/fossils (of a particular group) in an area	I can make a checklist	all collections by specific region	Metadata on collection level	Metadata on record level			
343	1	Scientist	know legal procedure and limits for loan of collection material and transport limits regarding dangerous goods for countries that host material	I can choose material for scientific work also by simplicity of loan procedure	database of legal and transport procedures and limits for loan of collection material by countries	Legal and policy framework				
344	1	Scientist	know the quality of the data provided	I can incorporate it into an analysis	it's up to date and certified (to some degree) by the institution publishing the data	Tools for data analysis				
345	1	Scientist	know where the specimens are kept in the collection	I can find them in the collection	to know the number of the boxes where the specimens are kept	Metadata on collection level	Tools for reporting & statistics			
346	1	Scientist	measure the geological or palaeontological specimen	I can compare the measurements and use them in publications	a scale/ruler in the digital image or a measurement tool for the image	Tools for identification	Tools for data analysis			

347	1	Scientist	query when and where one or more fossil species or rocks/minerals etc. have been recorded, and their characteristics, and the institutions that archive specimens	I can use more geological/palaeontological specimens, or borrow collections from other institutes	Taxonomic fields or geological classification, geographic coordinates, date of collection	Metadata on collection level	Advanced search functionality			
348	1	Scientist	read untranscribed label data	I can add specimen details to the record	digital images of specimen labels	2D images	Label images			
349	1	Scientist	retrieve the licensing information of an image	I know if I can use the image I want to include in a paper	a reference to the image license	Tools for data discovery	Digital representation of specimens			
350	1	Scientist	to find type specimens (of a fossil species or a mineral)	I can verify and understand its characteristics	to be sure this name is understood uniformerly	Advanced search functionality	Data integration			
351	1	Scientist	verify the validity of a determination by inspecting the type material for that mineral / fossil	I can confirm the valid identity of a mineral / fossil	high quality images with different views (dorsal, ventral, detail shots etc...) and the metadata for the type specimens	2D images	3D images	Tools for identification		
352	1	Scientist	visit a collection and annotate additional information of specimens through an Unified Curation and Annotation System (UCAS)	I can capture information on geographical coordinates, locality, scientific name, accession number, collector name, and relevant measurements of specimens	a CMS independent annotation system	Annotation tools	Interoperability	Data integration		
353	1	Scientist	what kind of specimens you have in your collections	I can get specimens on loan	access to a loan portal	Advanced search functionality	Metadata on collection level	Metadata on record level		

354	2	Citizen Scientist	add label information to the specimen records	I can contribute to scientific data	access to the DiSSCo portal	Advanced search functionality	Annotation tools				
355	2	Citizen scientist	help characterizing the collection	I can enjoy	images of storage units without transcription	Metadata on collection level					
356	2	Citizen scientist	help transcribe specimens	I can enjoy	images of specimens without transcription	2D images	3D images	Annotation tools			
357	2	Citizen Scientist	keep track of my records and corresponding samples in a way that makes donating the specimens to a local museum painless to all parties	I can focus on creating reliable records	public, easily understandable data formats, tools for creating labels with unique IDs, and a public DB system that can link sample IDs to occurence data	Reference system & Standard lists	Interoperabili ty	Data integration			
358	2	Citizen scientist	know where a certain collector was on a certain day	I can help to transcribe a specimen	an existing transcription of a specimen that was collected around the same time by the same collector	Advanced search functionality	Metadata on record level				
359	2	Citizen Scientist	read (handwritten) untranscribed label data	I can add specimen details to the record	digital images of specimen labels	Label images	Tools for annotation				
360	2	Collection manager	analyse the uniqueness of the institutional collection	I can communicate its value	to compare my collection with data on collections from other institutes	Metadata on collection level	Metadata on record level				
361	2	Collection manager	be able to access information about storage conditions/status for specimens	I can optimise the storage of my own collections and plan for future acquisitions	access to storage/status information of specimens	Metadata on collection level					

362	2	Collection manager	check in which institutions certain collection categories are kept	I can forward this information to a collection holder, I can forward a collection on offer to an institute that is interested	details about geography and possibly wish lists for certain specimens	Metadata on collection level				
363	2	Collection manager	connect a researcher to colleagues	they can examine more collections	to know which institute holds specific collections	Metadata on collection level	Data integration			
364	2	Collection manager	encourage remote curation of my collection through expert annotation	I can improve the curation and value of the collection	to receive and be able to easily incorporate annotated data	Tools for annotation				
365	2	Collection manager	have the highest possible level of data security	I can rest assured that nobody hacks the system, illegally modifies or extracts data	strict focus on data security during the setup of DiSSCo	Data security				
366	2	Collection manager	know which users are interested in which data	I can meet the needs of as many users as possible	information who the users are (e.g. citizens, scientists), where they are based (e.g. country, type of institution) and which data they are interested in (pictures, specific data categories e.g. vernacular names...)	Tools for reporting & statistics				
367	2	Collection manager	measure the use of collections via citations	I can understand the use of the collection and give evidence of its importance	to track specimen identifiers and their citation	Tools for reporting & statistics				

368	2	Collection manager	profile my collection	I can show the importance of the collection	to be able to highlight my institutional collection within the DiSSCo collection	Metadata on collection level	Tools for reporting & statistics			
369	2	Collection Manager	start a digitizing project	I can digitize a certain group of my collection, can do this internationally because of funding	to know where else there are collections of this group	Metadata on collection level				
370	2	Collection manager	understand the needs of researchers	I can make information available useful, and develop collections effectively	to know what researchers need	Metadata on collection level	Tools for clustering requests			
371	2	Collection manager, Director, Administrator	know the situation with collection amounts	I can plan ahead for future storage needs	to know existing amounts of collections, and amount of new material coming in	Tools for reporting & statistics				
372	2	Curator	add annotated information from an Unified Curation and Annotation System (UCAS) to my collection management system (CMS)	I can update infomation on my specimens in my CMS	interoperability between my CMS and UCAS	Interoperability	Tools for uploading	Annotation tools		
373	2	Curator	annotate all aspects of records, suggest improvements, and record logical connections between records	I can provide duplicate-free, reliable, well-documented data to end users	an record-level annotation and communication system spanning across institutes	Annotation tools	Tools for data analysis			

374	2	Curator	annotate digital specimens with updated determinations	I can improve the curation of the collection	to be able to annotate a digital specimen and pass that to the curating institute	Annotation tools	Images	Interoperability		
375	2	Curator	answer multiple requests on a specified group of rocks, minerals / collector / geographic area	I can follow conversation about a request	communication thread by rock type etc. / collector / geographic area	Advanced search functionality	Tools for clustering requests			
376	2	Curator	answer requests for specified objects	I can search the collection and pull material	to receive requests including all rock/mineral names involved	Metadata on record level				
377	2	Curator	attach and deliver geochemical data to rock and mineral specimen records	I can retrieve specimens based on their geochemical signature	a portal compatible or similar to the Earth Chem portal funded by NSF http://www.earthchem.org/	Tools for data discovery	Biochemical or geochemical data	Data integration	Tools for uploading	
378	2	Curator	be recognised as contributor	I can apply for funding to digitise institutional collections	contribution indicators (as contributor)	Tools for reporting & statistics				
379	2	Curator	check that the transcribed label data corresponds to the actual label information	I can confirm collection details	digital images of specimen labels	Label images				
380	2	Curator	compare an unidentified specimen with identified specimens to determine their identity	I can identify the specimen	digital images of specimens	2D images	Annotation tools	Tools for identification		
381	2	Curator	cross-check data between specimens collected by the same collector on the same day	I can confirm that all specimens have similar geographical coordinates, or correct coordinates where necessary	to select all DiSSCo records by collector and date	Metadata on record level	Tools for geo-referencing	Advanced search functionality		

382	2	Curator	curate a digital specimen (as it enters the DiSSCo data infrastructure)	my collection management system (CMS) has curated specimens	direct access to my digital specimens from the DiSSCo infrastructure	Tools for data discovery	Advanced search functionality	Annotation tools			
383	2	Curator	discover the type status of specimens	I know how many type specimens are in the institutional collection	digitised information on the description of species	Advanced search functionality	Data integration				
384	2	Curator	enrich my collection with reliable annotation from specialists anywhere in the world	I can increase the value of my collection	a quality / reliability rating of annotators	Annotation tools					
385	2	Curator	images (old or modern photos or drawings of the complete specimen or of any microscopic technique applied on it)	I can make the accurate identification	These images are extracted from the original publications or requested from their authors and stored	Tools for identification	2D images	Interoperabili ty	Data integration		
386	2	curator	monitor and in specific cases restrict access to geographical coordinates of collection sites	I can stop ruthless exploitation of fossils, certain minerals or sensitive sites	a possibility to personally evaluate every request for a combination of certain data categories, and the possibility to modify the answer to a data request	Tools for limiting access to data	Advanced search functionality				
387	2	Curator	publish my data online	I can increase the value of the collection	a user friendly collection management system (CMS)	Tools for uploading	Advanced search functionality				
388	2	Curator	read untranscribed label data	I can add specimen details to the record	digital images of specimen labels	Label images	Annotation tools				

389	2	curator	relate catalogue numbers of material in my collection to published scientific papers where they have been used	I can estimate and present scientific value of my collection	database with catalogue numbers of the specimens and the references of all scientific papers where they have been used	Data integration	Tools for reporting & statistics			
390	2	Curator/collections manager	increase the collections visibility for the general public	I can motivate amateurs/citizen scientists based on the value of diversity	specialized personnel to present parts of the collection in a story telling, yet scientifically sound, manner	Metadata on collection level				
391	2	Digitisation officer	ensure digitisation serves research needs	I can make effective use of resources	to know what researchers require	Advanced search functionality	Tools for clustering requests			
392	2	Digitisation officer	produce digital specimens from a digitisation line	I can store a specimen in my collection management system (CMS)	an automated workflow minting persistent identifiers (PIDs)	2D images	Label images	Interoperability	Tools for uploading	
393	2	Digitization officer	link specimen & label images the corresponding occurence data	I can make the images as publicly visible and usable as possible	a (semi)automatic label data extraction & verification system	Data integration	Label images			
394	2	Director	hire a curator with knowledge of specific groups	I can be sure they have a background that includes knowledge of the main collection	collection types, importance of collection gauged by size, scope, and time period of collection	Metadata on collection level				

395	2	Director	know how much my institution's collections is used and for what	I can argue for the importance of my institution's collection	to be able to extract information from DiSSCo based on # of views/downloads/annotations etc. of my collection	Tools for reporting & statistics	Legal and policy framework	Tools for downloading data/metadata	Metadata on collection level	Annotation tools
396	2	Historian	find information on the history of objects and collections	I can study the historical context of collections and objects	historical data, like previous owners, links with other objects, data of arrival in collection, previous ownership etc.	Metadata on collection level	Metadata on record level			
397	2	Scientist	compare an unidentified specimen with identified specimens to determine their identity	I can identify the specimen	digital images of specimens	2D images	Annotation tools	Tools for identification		
398	2	Scientist	correct an identification and add an annotation	I can file the specimen under the correct taxonomic name	digital images and an annotation system	Annotation tools	Tools for identification			
399	2	Scientist	cross-check data between specimens collected by the same collector on the same day	I can confirm that all specimens have similar geographical coordinates, or correct coordinates where necessary	to select all DiSSCo records by collector and date	Advanced search functionality	Metadata on record level	Tools for geo-referencing		
400	2	Scientist	cross-check data between specimens of the same rock/mineral etc.	I can flag outliers and correct record data where necessary	to select all DiSSCo records from a certain rock/mineral etc.	Advanced search functionality	Annotation tools			
401	2	Scientist	extract handwriting samples of a collector	I can verify collection localities and collection dates of specimens of a collector	to select all digitized labels from a specific collector	Advanced search functionality	Label images			

402	3	Automatic identification systems developer	know which collections are available to use as a reference (training data set)	I can train my algorithms for automatic identification	collections of target group (validated)	Metadata on collection level	Tools for identification			
403	3	IT support	build and provide solutions and related services	I can provide services to curators so that they can work better and easier with their collections at less financial costs	information on volumes, locality data, physical storage volumes, plus an insight on what is digitally represented and what is not	Tools for reporting & statistics	Metadata on collection level	Tools for data discovery		
404	3	Software developer	create new usages with the data and ways to add to the data, through apps or web interaction	data is more accessible to the masses and different collections can be, for instance, cross-referenced. At the same time additional data can be added and fed back into the core databases. Geographic location will be involved as every man has GPS access today. The vantage point to access these 'big data' sources could be educational, entertaining, medical, historical and natural sciences	Scope: Collection level, details: Specimen level	Data integration	Metadata on collection level	Advanced search functionality	Metadata on record level	Annotation tools
405	3	Solution provider	tap into the vast market of digital storage solutions for digital natural collections	I can sell my services and consult	predictable numbers on collection type, volume and progress in digitization	Metadata on collection level	Tools for reporting & statistics			

406	4	Association	to gather information to have overall figures representative of partners' state-of-the-art	we can showcase the relevance of the collections to policy makers and attract funds	high-level figures that feature the collections as a whole	Metadata on collection level	Legal and policy framework	Tools for reporting & statistics		
407	4	Collection manager, Director, Administrator	know what the situation is regarding collection size	I can plan for new space/ storage needs	I need to know existing size of collections, and amount of new material coming in. Also, I need to know the status/ condition (e.g. wet, dry) of existing material and collection health information	Tools for reporting & statistics	Metadata on collection level			
408	4	Director	know the extent of the use of the organisation's collections across societal sectors	I can take informed executive decisions related to future investments (both collections and HRs)	metrics of the use of collections	Tools for reporting & statistics	Metadata on collection level	Legal and policy framework		
409	4	Director	to be able to underline the importance of Earth sciences and of scientific collections in understanding it	I can help policy makers understand the consequences of their policies	details of collections and related research	Tools for reporting & statistics	Metadata on collection level	Legal and policy framework		
410	4	Director	understand the relationship of our own collection to the collections of other institutions in our country	I can explain the scientifc value and unicity of our collection to policy makers	information on collection type, size and taxonomic breadth of other institutes	Tools for reporting & statistics	Metadata on collection level			

411	4	Director / administrator	know what makes our collection unique	I can effectively advertise/highlight the collections to improve usage	Collection types, with size, locality scope, time, taxonomic scope, important collectors	Metadata on collection level	Tools for reporting & statistics	Advanced search functionality	Metadata on record level	Data integration
412	4	Policy maker	find and reuse digital specimens from DiSSCo	I can confirm the presence of a rock/mineral etc. for legal purposes	fast access to the DiSSCo infrastructure	Advanced search funcionality	Legal and policy framework			
413	4	Policy maker	find information on the contribution of DiSSCo to the international environmental policy agenda	I can justify national level investments in its operation	evidence that DiSSCo's activities align with (e.g.) Sustainable Development Goals	Legal and policy framework	Tools for reporting & statistics	Data integration		
414	4	Policy maker	know how well sampled my country is	I can fund future biodiversity exploration and research	an extractable list of specimens to understand which parts of my country are poorly known on an easy to see map	Metadata on collection level	Tools for reporting & statistics			
415	4	Policy maker	know the use of the collections by other domains as a key indicator of its impact	I can distribute resources and allocate them in alignment to the strategic priorities of the government that I represent	information on access to the collections, virtually and physically, from different types of users	Tools for reporting & statistics	Legal and policy framework			
416	4	Policy maker	understand the value of DiSSCo	I can justify the national level investments in its operations	access to impact stories and/or assessments	Tools for reporting & statistics	Metadata on collection level			
417	5	Curious person	learn about the rocks/minerals/fossils that might be in my environment	I can improve my knowledge on geology / mineralogy / palaeontology	scientific names, common names, geographic coordinates, characteristics, images	Metadata on record level	Data integration			

418	5	Exhibitions maker	be able to find the location of specimens and information about species	I can design and realise interesting and accurate exhibitions and if required, loan specimens from otgher institutions	information about the location and status of specimens in museum collections	Metadata on collection level	Metadata on record level			
419	5	Science communicatio n officer	know about the research being carried out in natural history collections	I can organize a dissemination event	information on research topics and the people behind them	Data integration				
420	5	Student	be able to identify the species in scientific field trips, without visiting the relative collections of NHMs	I save time and money, by avoiding the physical travelling to those NHMs	to have access to the NHMs' digitalised species' identities (morphometric characters, names, distributions, etc.)	Tools for identification	Advanced search functionality	2D images	Morphologica l data	Distribution data
421	5	Student	check my identification of a geological/palaeontoogical specimen	I can test my knowledge about rocks/minerals/fossils and and learn more about a group I am studying	image/characteristics, links to papers, collection metadata	Tools for identification	Images	Metadata on collection level	Data integration	
422	5	Student	find a high resolution image	I can perform morphological analysis as part of my coursework	a reference to the higher resolution image and the clear procedure to retrieve it	Tools for identification	2D image	3D image	Tools for downloading data/metadat a	
423	5	Student	find the referencing information for an image	I can cite the source in a written report which includes the image	a reference to the citation information for the image	Metadata on record level	Data integration			
424	5	Teacher	be able to provide accurate scientific information about rocks/minerals and specimens	I can educate students/the public about the natural world	easily accessible and up-to-date information about the specimens in museum collections	Metadata on record level	Advanced search functionality			

425	5	Teacher	find a high resolution 3D model	I can use a hologram projector for inspection of the specimen in class	a reference to the available 3D models	Advanced search funcionality	3D images	Data integration		
426	5	Teacher	find out if there is a printable version of a 3D model	I can create a 3D physical model for study in class	a reference to the available 3D models	Advanced search funcionality	3D images	Data integration		
427	5	Teacher	find a high resolution 3D model	I can develop a folding forms for students (the broader public)	a reference to the available 3D models	Advanced search funcionality	3D images	Data integration		
428	5	Teacher	get microfossil microscope slides	I can use them for teaching at highschools or universities	(sets of) microfossil microscope slides (physical objects) plus metadata	Metadata on record level	Physical access			
429	6	Developer	find a 360 degree view or a 3D model of a specimen	I can use it in the creation of interactive content for use with augmented reality educational software	a reference to the available 3D models	3D images	Advanced search functionality			
430	6	Digitisation officer	produce digital specimens from a digitisation line	I can upload the images to a customer's CMS	an automated workflow minting PIDs	Tools for uploading	Data integration			
431	6	Publisher	include a digital image from a collection in a scientific paper	I can illustrate the publication	access to the digital images and copy rights to use the image	Legal and policy framework	2D images	Data integration		

432	6	Software developer	develop new usages of the data and ways to add information to the data, through apps or web interactions	I can make the data easier accessible to the general public and facilitate that different collections can be, for instance, cross-referenced. At the same time the additional data can updated in the core databases. This includes addition of geographic locality data as most users hold an handheld GPS device. The vantage point to access these 'big data' sources could be educational, entertaining, medical, historical and natural sciences.	detailed collection level information	Metadata on collection level	Data integration			
433	6	Solution provider	build and provide solutions and related services	the curators and scientists can work better and easier with their collections at less financial costs	information on volumes, locations and physical sizes plus an insight on what is digitally represented and what not. Even better would be if there is an institutions priority as to what needs to be digital first.	Tools for reporting & statistics	Metadata on collection level	Tools for data discovery		

434	6	Solution provider	tap into the vast market of digital storage solutions for digital natural collections	I can sell my services and consultancies	predictable numbers on collection type, volume and progress in digitization	Metadata on collection level	Tools for reporting & statistics	Tools for data discovery		
435	6	Artist	use information on historical ecosystems (e.g. 3D scans of fossils)	I can use it during the creation of an historical art installation	2D or 3D scans of fossils	3D images	2D images	Data integration		
436	6	Film maker or artist	film the (micro)fossil collections and interview scientists/curators	I can get footage for movies or art projects	"stories" related to the collections (e.g. historically important collections or somewhat fascinating material/species)	Metadata on collection level	Metadata on record level	Data integration		
437	6	Curator or researcher working in a micropaleontology collection	use my expertise about certain fossil groups (e.g. forminifera, dinoflagellates, etc.)	I can work as a consultant for companies (e.g. in the oil industry)	collection data for certain fossil groups	Metadata on collection level	Metadata on record level			
438	6	Curator or researcher working in a collection	use my expertise about certain habitats, sediments, soil types, and/or organisms	I can work as an expert or consultant in forensics and support police or law enforcement agencies (e.g. to identify the locality of soil traces or determine fresh vs. salt water, etc.; forensic geology/limnology)	the collection as reference set of materials, to search for relevant collection objects	Metadata on collection level	Metadata on record level			
439	6	Administrator or laboratory manager	rent out the lab facilities to external commercial users	I can make full use of the capacity of the laboratory and optimise degree of utilization	infrastructure rather than collection	Physical access				

440	7	Journalist	link to primary source data (scientific literature, museum collections databases, etc.)	my readers can learn more about the topic of an article	collections database records	Metadata on record level	Data integration			
441	7	Journalist	film the (micro)fossil collections and interview scientists/curators	I can get footage for documentaries	"stories" related to the collections (e.g. historically important collections or somewhat fascinating material/species)	Data integration	Metadata on collection level			
442	7	Public relations officer / Press officer	visualize what is currently stored in the collections and how it has developed and is developing over time using maps or other graphical representations	I can use it for self-marketing or for public information	to be able to sum up certain data categories from the collections and to export these data to be able to use them in special data visualization tools	Tools for reporting & statistics	Tools for downloading data/metadata	Tools for data visualization		
443	7	Marketing officer	have infographics presenting information and data on the collection incl. geographical or temporal visualization of numbers and sums, connections, relations, and correlations	I can offer these infographics to newspapers or use them for self-marketing	to be able to sum up certain data categories from the collections and to export these data to be able to use them in special data visualization tools	Tools for reporting & statistics	Tools for downloading data/metadata	Tools for data visualization		

CAT (Category based on ICEDIG T6.2)

1. Research (academic & non-academic, including citizen science)
2. Collection management
3. Technical support (IT & IM)
4. Policy (institutional, national & international)
5. Education (academic & non-academic)
6. Industry
7. External (media & empowerment initiatives)

Appendix 3. SYNTHESYS Transnational Access Analysis

In addition to surveys and approaching (potential) new users of the DiSSCo research infrastructure, NHMUK analysed the SYNTHESYS Transnational Access (TA) programme. As of September 2020, the TA dataset covers over 4,450 funded visits totalling over 54,000 researcher days to 26 collection-holding institutions in 14 countries with c. 12,300 self-reported research outputs. The full report includes background information and context, methodology (database review, standardisation, schema analysis and modelling, analysis database construction, data cleaning and enhancement), analysis (preparation of a publishable analysis database with removal of personally identified information (PII), DOI matching). The report will be submitted as a formal publication before June 2021. A summary of the results for demographics, functional demands on collections and facilities, and socio-economic impact is given below.

The SYNTHESYS Transnational Access data was harder to analyse than we had anticipated. When the access database was originally created it was not designed to analyse outputs. There was no verification of outputs by an administrator so the data quality is variable and contains user errors. Manual verification of the entire output dataset is time consuming – designing future systems that track the research outputs associated with collections or specific facilities should utilise automated validation or using controlled vocabularies. This would make subsequent analysis easier.

Future analysis could study the authors and their collaborators to understand their research fields and backgrounds. Data mining acknowledgements to study co-funding sources would potentially give an insight into cross-disciplinary work. Studying publications with the highest downstream citations would give an alternative metric for understanding impact of research but comes with its own limitations.

In terms of functional requirements for facility and collection access, further work could be usefully carried out on the relationships between collection access requirements and facility access requirements. One promising area of investigation would be to analyse any trends between collection type and analysis facility usage: does more than one installation per project reflect a dependency between collection access and availability of analysis facilities, or is the cause of this overlap driven by more pragmatic factors around availability of access to collections/facilities during a SYNTHESYS-funded visit? These kinds of questions may be useful for DiSSCo when trying to anticipate the needs of its future users.

The absence of a controlled vocabulary around the types of collection users wish to access makes it difficult to get more granular data around demand for, and usage of, these collections. Time permitting, more detail could potentially be extracted by running a Named Entity Recognition algorithm over unstructured data fields that are likely to contain useful information, e.g., project title and anticipated research benefits.

For future tracking of outputs across DiSSCo partners and in SYNTHESYS successors we recommend using a more sophisticated system that would support text mining and automated analysis. Tracking outputs from facility and collection usage is an important metric but poorly supported by the systems and data we currently have to hand.

Demographics

As of late 2019, 64% of applications submitted to the SYNTHESYS TA programme were made by men, 36% by women.

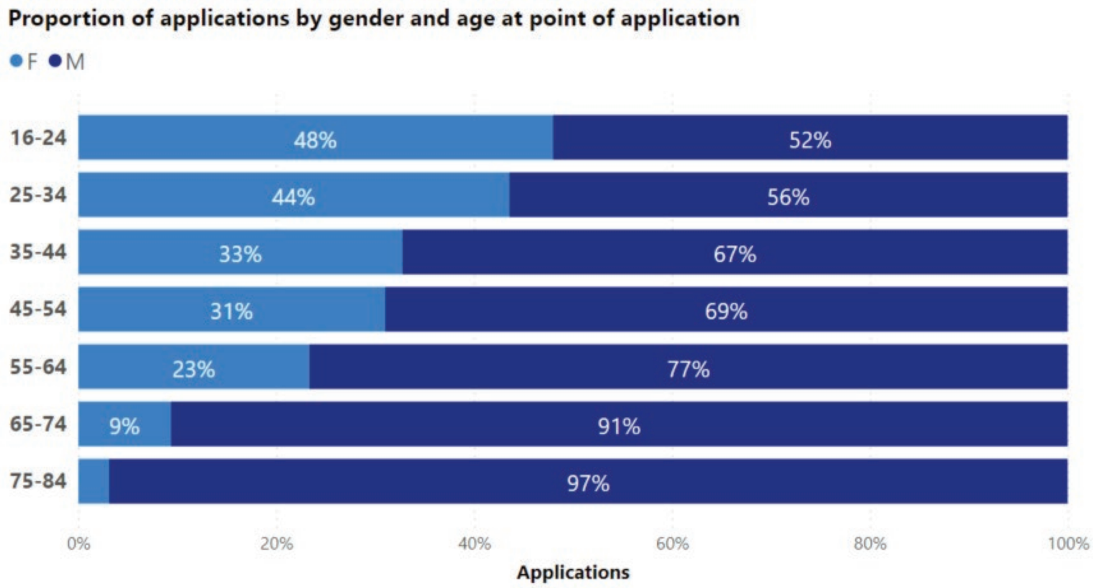

Figure A3.1. Proportion of applications by gender and age at point of application (F - female, M - male).

Women were most likely to apply during the postgraduate stage of their career (44% of all applications by women), whereas the dominant career stage recorded by male applicants was 'experienced' at 37% of the total (see Figure A3.1). Overall, 34.5% of applications were made by postgraduates, 31% by postdocs, and 31% by 'experienced' career-stage researchers. Technical applicants comprised 1.5% of overall applicants and undergraduates 2.5%. The most common age bracket for applicants at the point of application is 25-34 years of age (43.6% of all applications submitted), followed by 35-44 years (27.4%). The prevalence of younger, earlier-career applicants in the TA scheme is not surprising: established researchers are more likely to have additional avenues of funding available for a research trip. The drop-off in numbers of female applicants after their early 30s seems likely to reflect a paucity of time, rather than funding (Cech & Blair-Loy 2019).

Functional demands

The functional requirements for all applications were investigated in order to derive a more accurate overview of functional demands for institutional collections and services independent of TA Programme's administrative practices, which exist in part to ensure that demand for a particular category of service or collection does not overwhelm available host capacity.

At the national level, institutions in the UK received the most access requests through the SYNTHESYS TA scheme (30% of total), followed by France (11%) and the Netherlands (10%). Approximately 60% of all access requests are made to collections and the remaining 40% are targeted towards analytical facilities.

The majority (58%) of applicants requested access to more than one collection and/or analysis facilities within a single project application. Request distribution over the different access categories has stayed consistent over time (see chart below, Figure A3.2). SYNTHESYS Round 4 data is incomplete because it is still ongoing, so the increase in imaging requests visible in this round may not be sustained throughout.

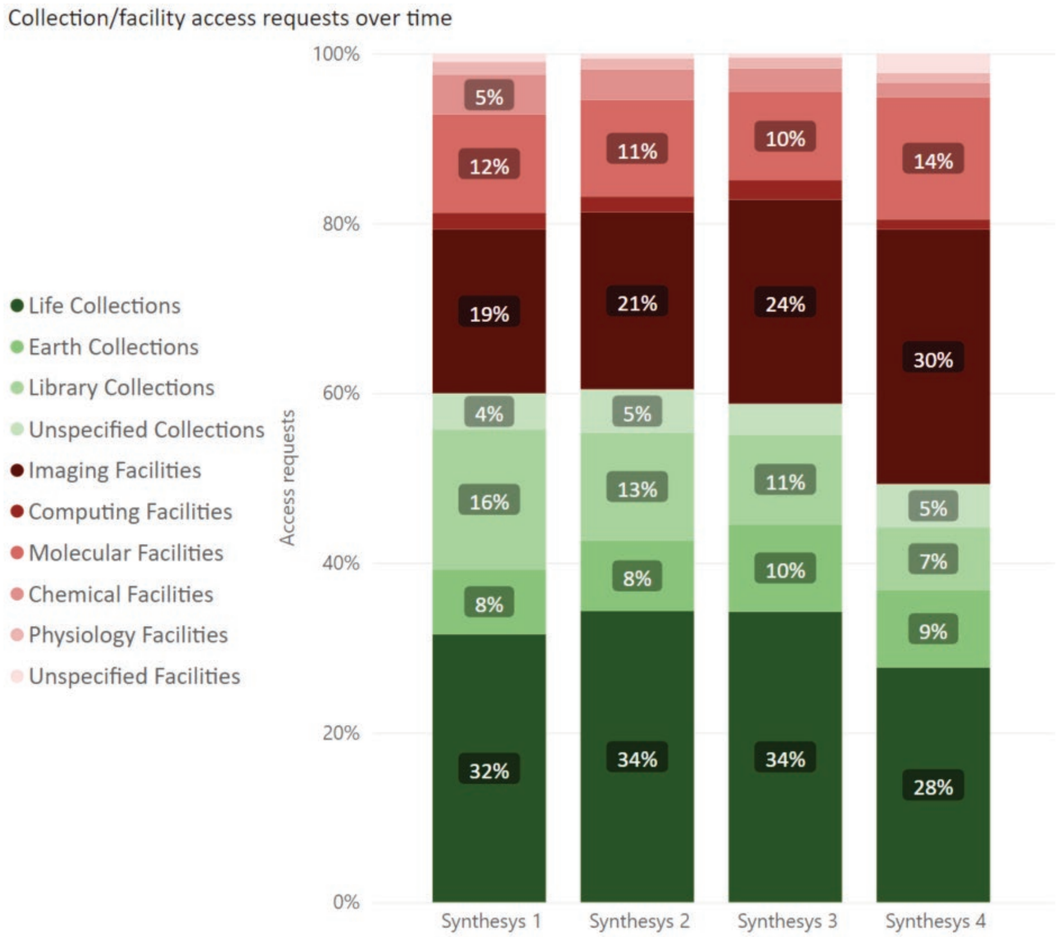

Figure A3.2. Chart showing distribution of request count by category of host facility or collection. Round 1 (n = 8.5k), Round 2 (n = 5k), Round 3 (n = 6.2k), Round 4 (Incomplete: n = 2.8k).

User discipline diversity and socio-economic importance

CrossRef records were found for 2,780 articles (22.6%) of the 12,280 self-reported research outputs provided by users after removing duplicates. Of these, 2,431 were identified by searching for the title and authors using the CrossRef API's search functionality, 672 (20.7%) had a user-provided DOI, and the remainder were identified from DOIs extracted from other user-provided data.

These CrossRef article records were then checked for potential discipline diversity and socio-economic impact using two approaches:

1. Automated journal subject tagging – looking for atypical subject tags that were neither life sciences or earth sciences
2. Manual title checks - looking for papers of more immediate societal relevance

From the automated journal subject tags we counted the number of non-Life/Earth Sciences research outputs in each subject tag category (see Figure A3.3). The majority fell into general medicine with a fairly broad distribution across 28 other categories. While this was a useful summary it did not provide enough information to make any judgements on socioeconomic importance.

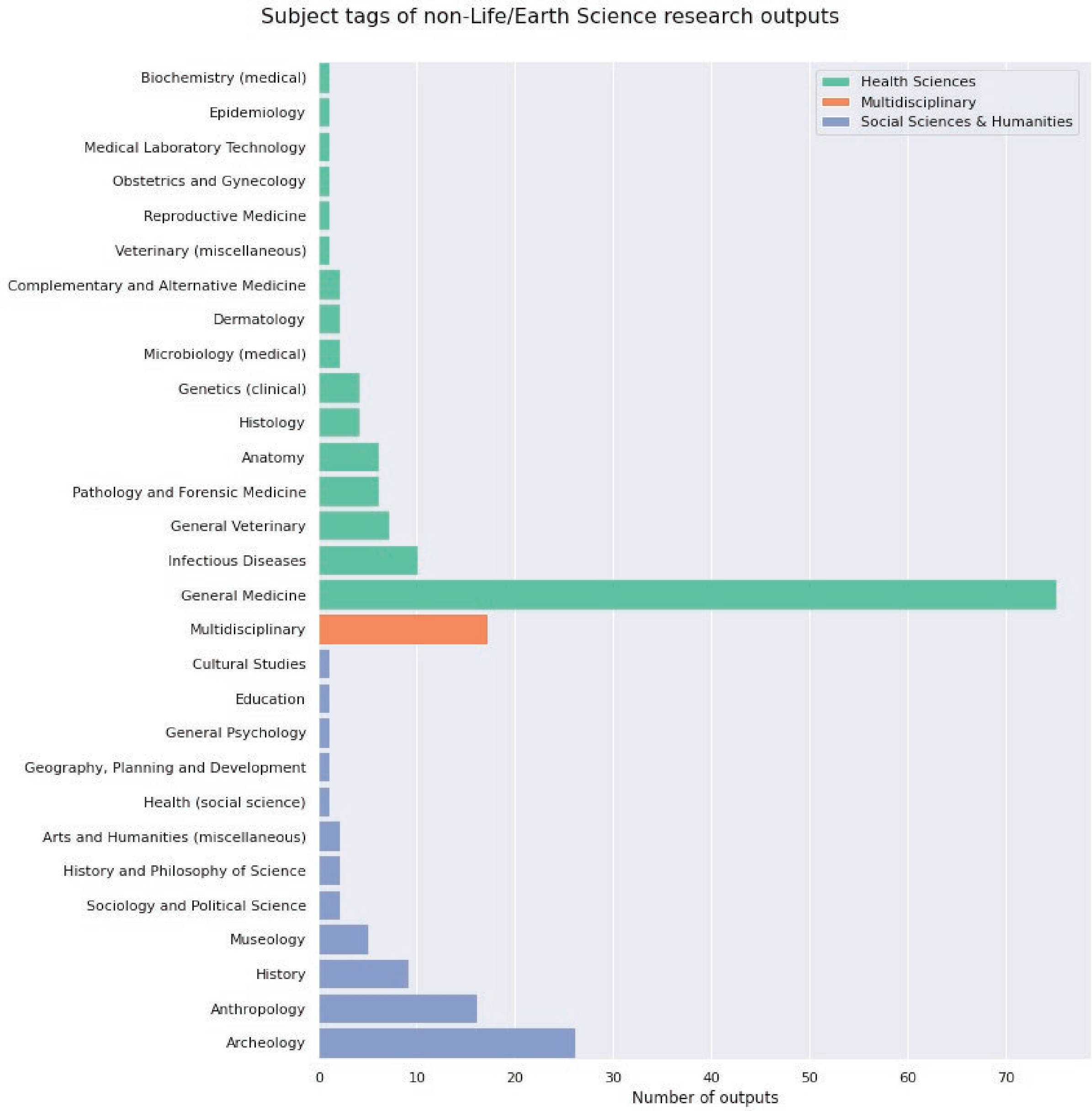

Figure A3.3. Subject tags of non-Life/Earth Sciences research outputs.

Manual title checks were the most reliable way of checking whether a research output mapped to one of the seven H2020 Societal Challenges which we were using as a proxy for socioeconomic importance. We manually checked each of the 2,780 publication titles and verified if the paper cited SYNTHESYS in either the acknowledgements or funding metadata. This resulted in 199 outputs that mapped to a societal challenge of which 49 acknowledged SYNTHESYS.

The majority (37) mapped to the "Climate action, environment, resource efficiency and raw materials" and depending on how strictly you consider the "environment" component of the challenge then many more of the 2,780 could be assigned here. Six outputs were assigned to the "Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the Bioeconomy" with the final six evenly distributed into "Health, demographic change and wellbeing", "Europe in a changing world – inclusive, innovative and reflective societies" and "Secure societies – protecting freedom and security of Europe and its citizens."

Appendix 4. Target Groups for additional surveys and interviews. Given is the target group or contact point and a use case category where surveys contributed to or might contribute in future.

Target Group / Contact Point	**Use Case Category**
Geo- and Paleo colleagues from DPP Partner Institutions	1. Research
Geo.X (Research Network for Geosciences)	6. Industry
TDWG Palaeo & Earth Science Interest Group	several
PR team from DPP Partner Institutions	7. External
Mediasphere for Nature Network with a focus on particular partners	6. Industry
Educational sector from DPP Partner Institutions	5. Education
Partners from Synthesys+ NA5 Survey (Engaging with the private sector: Experience of institutes)	6. Industry
Geological Service / geological state agency	4. Policy
'Umbrella Organization for Geosciences' (*Dachverband der Geowissenschaften* , https://www.dvgeo.org/	several
Federal Institute for Geosciences and Natural Resources	4. Policy
(via personal contact)	
The German Mineralogical Society (via personal contact)	1. Research
	5. Education
The German Paleontological Society (via personal contact)	1. Research
	5. Education
Different Industries / Companies	6. Industry
(via personal contact)	
Film Academy	5. Education
(e.g. Konrad Wolf University Babelsberg)	7. External
Broadcasting Company	5. Education
	7. External
Lyme Regis Fossil Festival	5. Education
CETAF Earth Sciences group experts	1. Research