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 D1.1 Report on Life sciences use cases and user stories

Author(s)	Identifier of the author(s)
Fitzgerald, Heli	https://orcid.org/0000-0002-6754-6409
Juslén, Aino	https://orcid.org/0000-0001-9434-5250
von Mering, Sabine	https://orcid.org/0000-0003-2982-7792
Petersen, Mareike	https://orcid.org/0000-0001-8666-1931
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
Luomus - Finnish Museum of Natural History, Helsinki, Finland

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/xhxw-cb79 (D1.1.)

Publication year
2021

Related identifiers
https://doi.org/10.7479/17qp-ge55 (Data publication)
https://doi.org/10.34960/n3dk-ds60 (Deliverable D1.2)

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
Fitzgerald, H., Juslén, A., von Mering, S., Petersen, M., 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/xhxw-cb79

Abstract
This Deliverable for D1.1 from the Horizon 2020, INFRADEV-02-2019-2020 project DiSSCo Prepare reports the results of Task 1.1 "Analyse Life sciences use cases and user stories". The complemented corpus of Life 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. Based on recognized functional demands recommendations for the related ongoing Work Packages and DiSSCo services development in general are given.

Content keywords
scientific

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP1

Project output
Deliverable

Deliverable/milestone number
D1.1.

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
heli.fitzgerald@helsinki.fi

H2020-INFRADEV-2019-2
Grant Agreement No 871043

DiSSCo Prepare WP 1 task 1.1. – Deliverable: D1.1 Report on life sciences use cases and user stories

WP Lead: Henrik Enghoff (UCPH)

Task Lead: Aino Juslén (Luomus)

Authors: Heli Fitzgerald (Luomus), Aino Juslén (Luomus), Tea von Bonsdorff-Salminen (Luomus), Sabine von Mering (MfN), Mareike Petersen (MfN), 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), Frederik Berger (MfN), Jiří Frank (NM), Tina Loo (Naturalis), Patricia Mergen (Meise BG), Patrick Semal (RBINS)

Abstract

This Deliverable for D1.1 from the Horizon 2020, INFRADEV-02-2019-2020 project DiSSCo Prepare reports the results of Task 1.1 “Analyse Life sciences use cases and user stories”. The complemented corpus of Life 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. Based on recognized functional demands recommendations for the related ongoing Work Packages and DiSSCo services development in general are given.

Key words

Biology, collections, DiSSCo, life sciences, 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

Contents

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 WP1, 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. Few additional use cases were included from responses by scientists, colleagues and other stakeholder groups.

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. 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. Figure 1 and 2. summarizes the number of use cases collected per use case category for both Life sciences and Earth sciences.

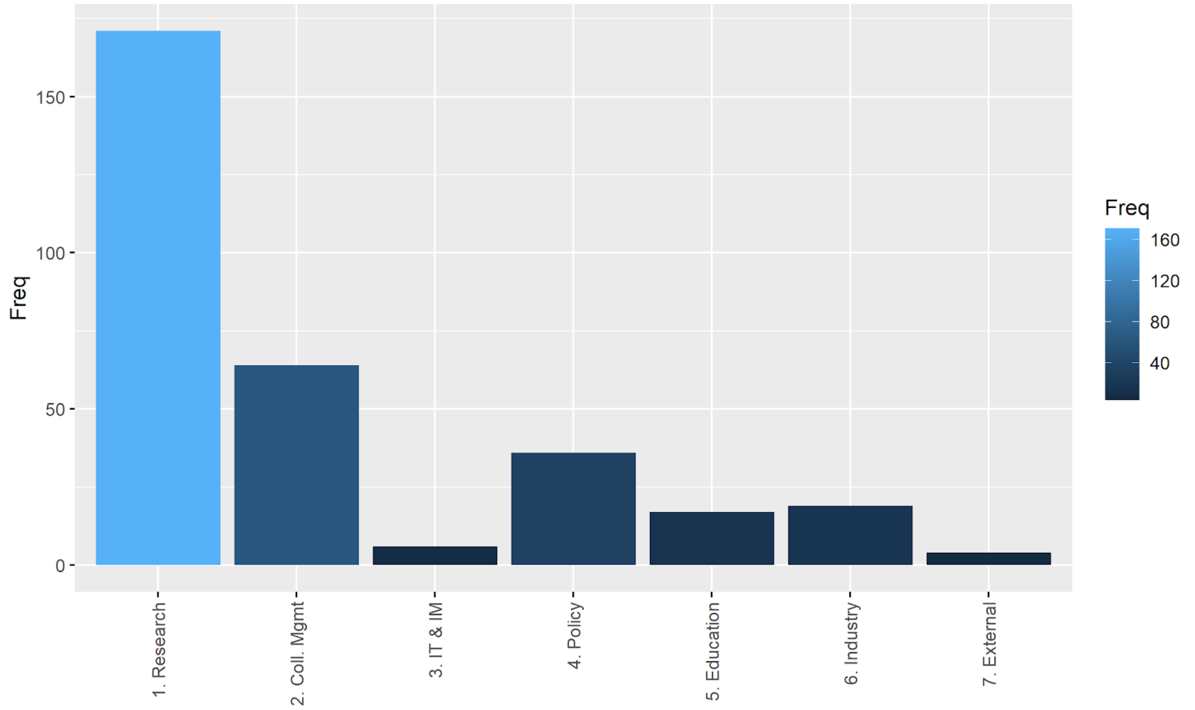

Figure 1. Number of use cases/user stories collected per category for Life sciences, total number of use cases n= 317

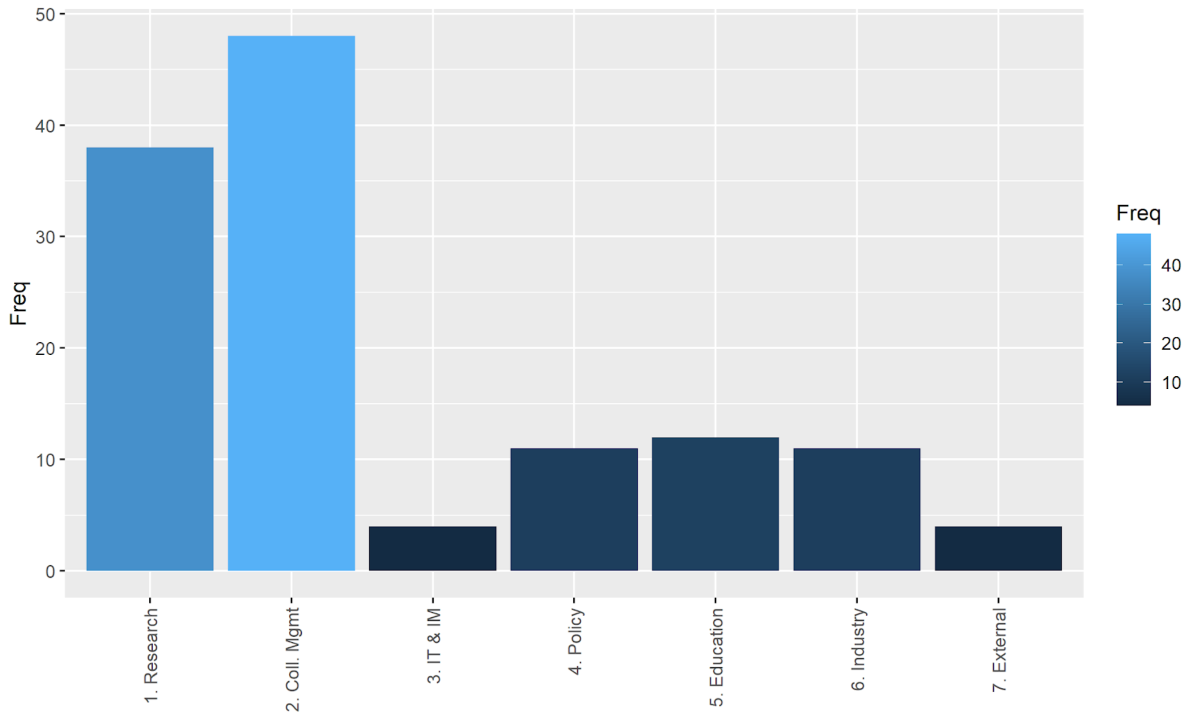

Figure 2. Number of use cases/user stories collected per category for Earth science, total number of use cases n= 128

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 linked with 11 categories and 24 subcategories. The most diverse categories were data and tools with six and 12 subcategories. Definitions for all categories and subcategories are given in Table 1.

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.

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 visualisation	Tools to visualise 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

Table 1. Functional demand categories/subcategories with definitions

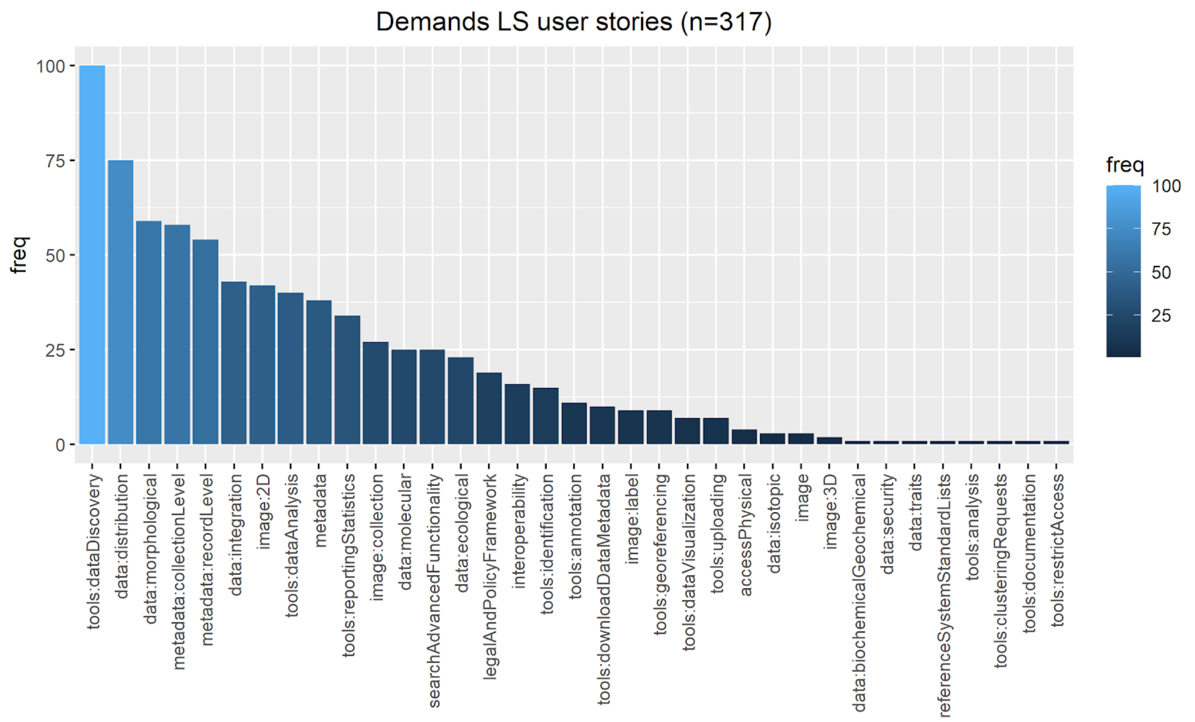

Figure 3. Number of times each of the 35 demands was mentioned in Life Sciences user stories

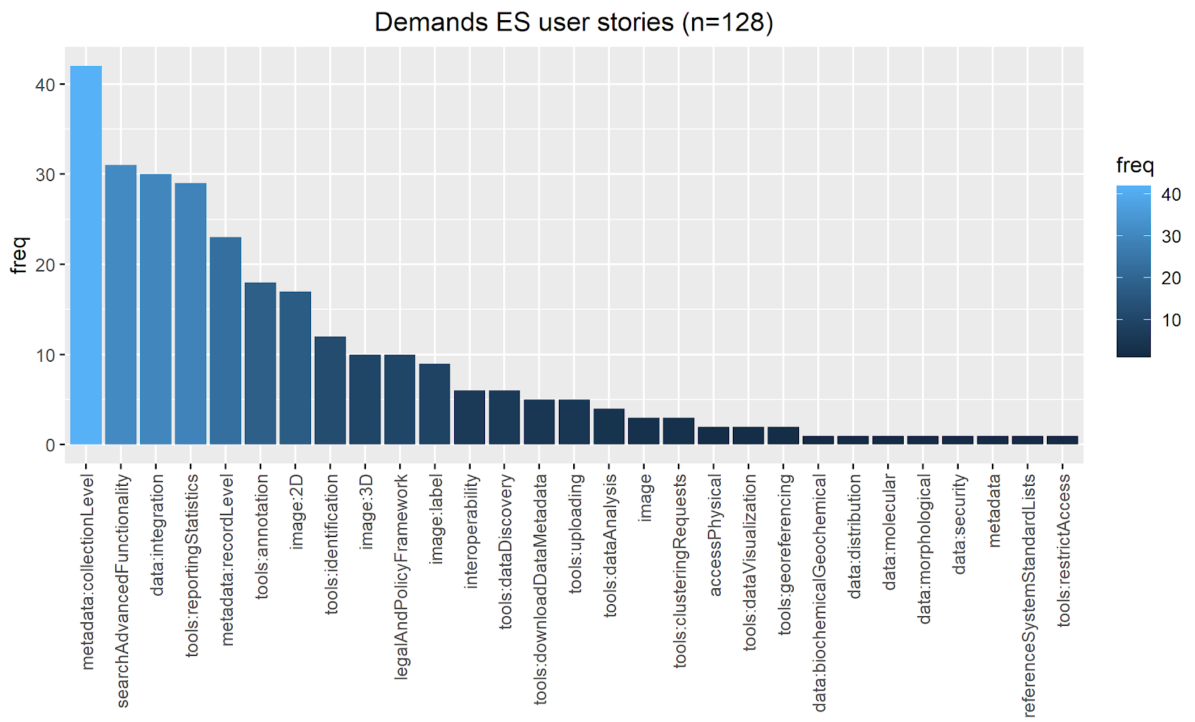

Figure 4. Number of times that each of the 29 demands was mentioned in Earth sciences (ES) user stories.

Analysing the user stories

The cluster analysis of the Life science use cases/user stories resulted in 13 different cluster groups, shown in different colors in Figure 5. These cluster groups correspond to use cases grouped together by having a combination of similar functional demand categories (shown in Figure 7.). The light blue cluster group example (Figure 6.), shows the grouping of use case IDs based on their functional demands. The clustering functional demands in this group are ‘Legal and policy framework’ and ‘Tools for reporting & statistics’. The full list of functional demands behind the user story IDs and cluster dendrogram can be found in Fitzgerald et al. (2021).

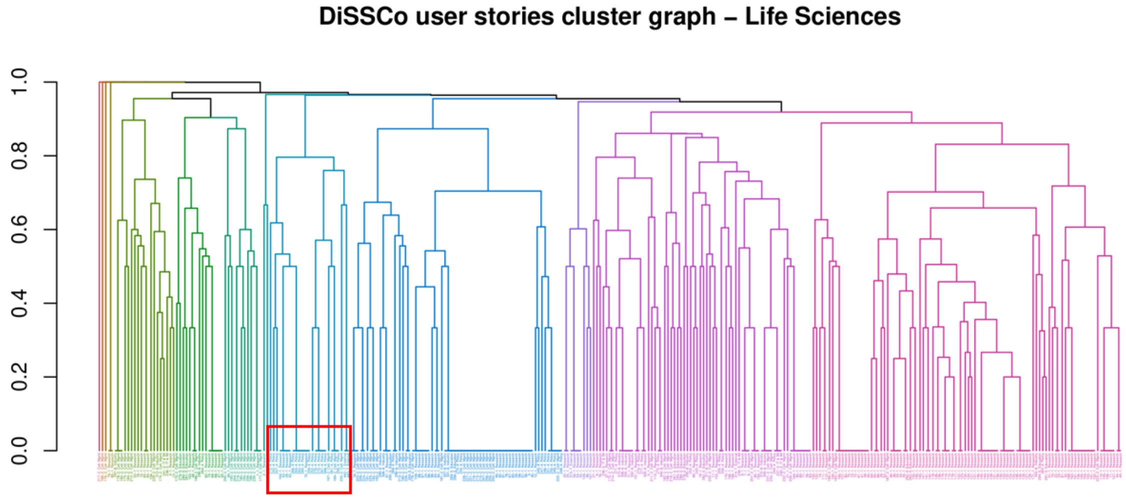

Figure 5. Cluster dendrogram for Life Sciences user stories (n=317). Different colors correspond to 13 different cluster groups.

Figure 6. Cluster group example (area within cluster dendrogram shown as red rectangle in Figure 3.). Use case IDs and user category included.

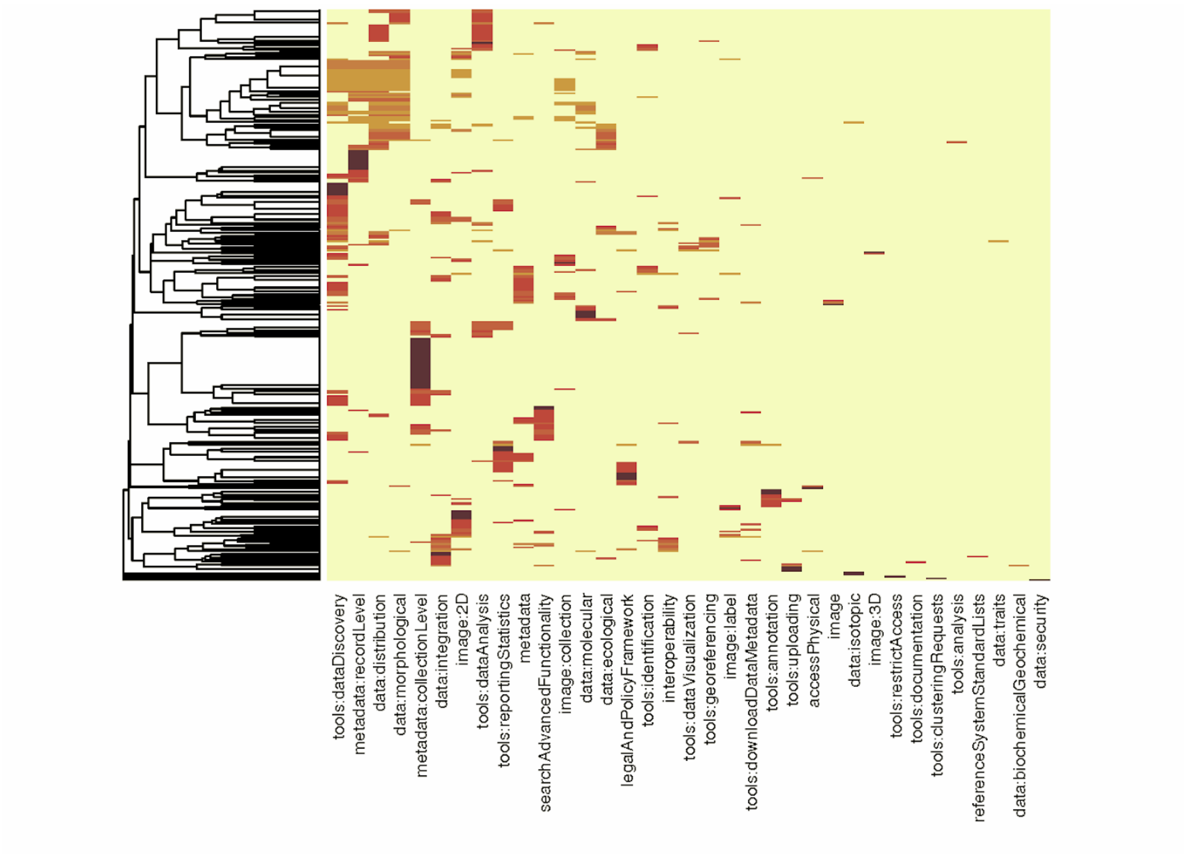

Figure 7. Heatmap of Life Sciences user stories (n=317). Left cluster diagram corresponds to Fig.3. The x-axis shows the 35 reported functional demands.

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.

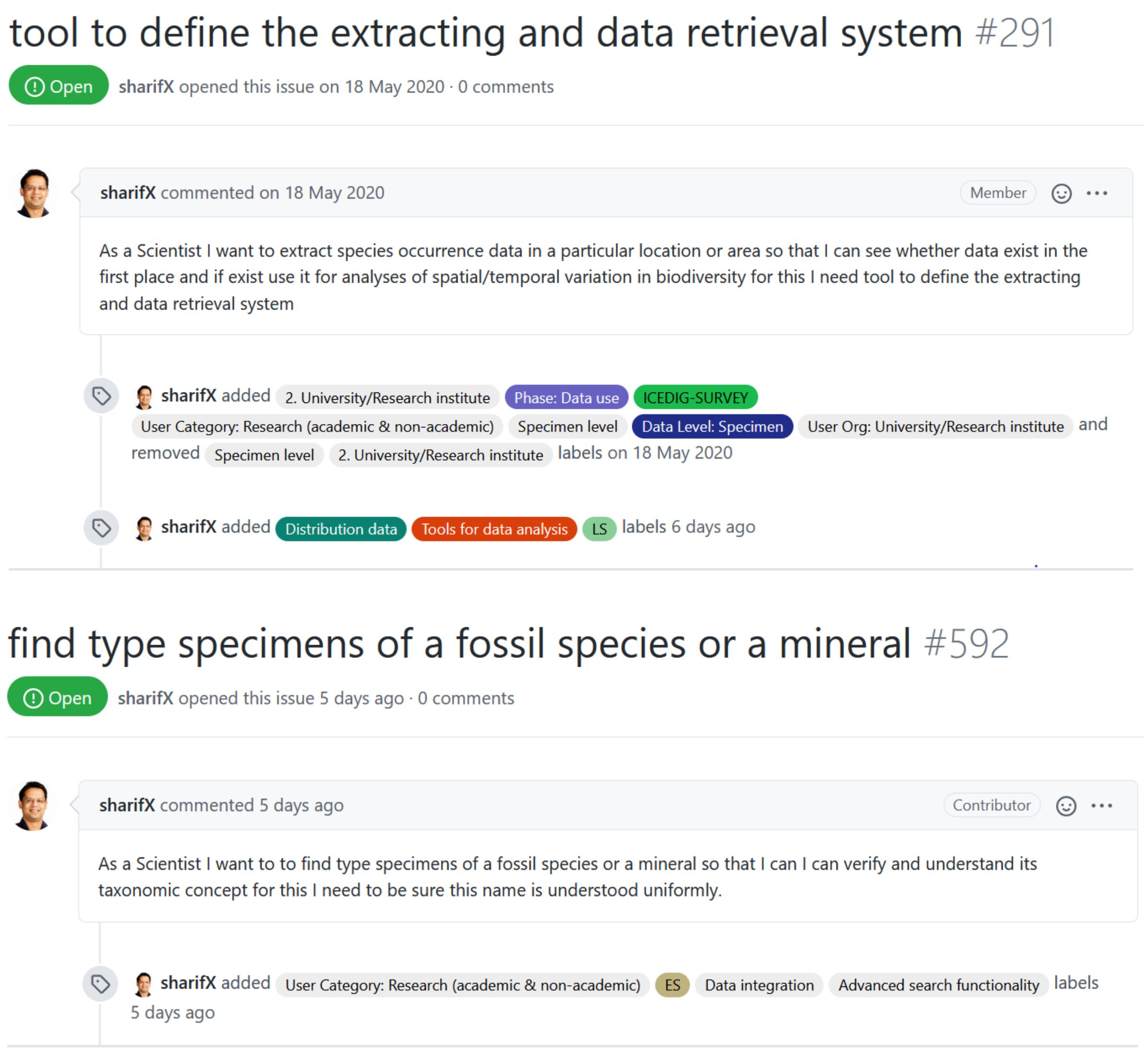

Figure 8. Screenshot of selected use case in GitHub repository shows tags for functional demands / use categories

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

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. Especially those which require more then only one functional demand and 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 (10): 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 Accessed ###

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

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(1):16. doi: 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 Life 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
1	1	Researcher, Scientist	visit a collection and annotate additional information of specimens through an Unified Curation and Annotation System (UCAS)	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				
2	1	Citizen Scientist	curate and add untranscribed labels	can contribute to the overall project, perhaps particularly on particular species groups	a curation interface	Physical access				
3	1	Researcher, Scientist	find out what is in the collection but not been digitised	know whether its available for me to use	a high level description of the collection	Metadata on collection level				
4	1	Researcher, Scientist	look at a specimen that has been sequenced	can confirm the identification before downloading the sequence from genbank	a link from genebank to the specimen and a fully databased record with a high quality image of the specimen	Interoperability	Data integration	Tools for identification		
5	1	Researcher, Scientist	examine a specimen that occurs in a GBIF record	can confirm the identification and the geographic co-ordinates	a link to the specimen from GBIF and a high quality image of the specimen and an image of the label data	Metadata	Interoperability	2D images	Label images	Tools for identification
6	1	Researcher, Scientist	download a collection of images with a resolution of 400 dpi in jpeg 2000 format	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 data discovery	Metadata	Images	Tools for downloading data/metadata	
7	1	Researcher, Scientist	find out if there are additional images available from a particular specimen	include different views of a specimen in a written report	a reference to related images for a particular specimen	Tools for data discovery	2D images	Data integration		
8	1	Researcher, Scientist	find out if there is an alternative image available with a resolution of 300 dpi and in png file format	select the appropriate image for including in a paper according to the publisher requirements	a reference to the alternative image formats available	Tools for data discovery	2D images	Data integration		
9	1	Researcher, Scientist	retrieve the licensing information of an image	know if I can use the image I want to include in a paper	a reference to the image license	Tools for data discovery	2D images	Data integration		
10	1	Researcher, Scientist	Measure the specimen	can compare the measurement differences between taxa/populations	a scale/ruler in the digital image or a measurement tool for the image	Tools for data analysis	2D images	Tools for identification		
11	1	Researcher, Scientist	study morphological variability of a taxa	can generate descriptions with extreme and median values	a tool extracting morphological information, and statisticaly represent this information	Tools for data analysis	Morphological data			
12	1	Researcher, Scientist	find all marker sequence linked to a taxa and obtain statistic on spatial and taxonomic distribution of all specimens	can build reference database	a tool that parse correctly taxonomic data, spatial information	Tools for data analysis	Data integration			

13	1	Researcher, Scientist	link specimens that belong to the same species according my taxonomic expertise (without a priori assigning species names)	can tag all specimens that belong to one species	a toolbox to link persistent identifiers (PIDS)	Data integration	Interoperability			
14	1	Researcher, Scientist	what kind of taxa and specimens you have in your collections	loan specimens	access to a loan portal	Tools for data discovery	Physical access	Metadata on record level		
15	1	Organologist	access to taxonomic backbone	investigate fault lines between biological sciences and organology. Organologists use biological sciences as examples for their taxonomies, classification systems, paradigms. Even the 'evolutionary theory' of musical instruments has much to learn from cladistics.	access to taxonomic backbone via integrated natural science collections data and archive resource	Data integration	Reference system & Standard lists			
16	1	Researcher, Scientist	know all species (of a particular group) in an area	can make a checklist	all collections by specific region	Tools for data discovery	Advanced search functionality			
17	1	Researcher, Scientist	extract spatial distribution data of a species	can use the data in a species distribution model (SDM)	all specimens from a certain species from DiSSCo with spatial coordinates	Tools for data discovery	Tools for geo-referencing	Distribution data		
18	1	Researcher, Scientist	extract all specimens with certain traits	can study the evolution of certain traits	all specimens from DiSSCo with a trait selected from Encyclopedia of Life (EOL) traitbank	Data integration	Ecological data			
19	1	Researcher, Scientist	extract a list of specimens associated in a particular area	Understand alpha and beta diversity turnover	all specimens geo-referenced	Metadata on record level	Tools for geo-referencing	Distribution data		
20	1	Researcher, Scientist	know if ther are cryptic species related with the species of my interest	can know related species	all the related information (molecular markers, articles,...)	Ecological data	Molecular data			
21	1	Researcher, Scientist	discover undescribed species in the collection	can describe new species	an AI (Artificial Intelligence) toolbox trained on described and correctly identified specimens to filter specimens that cannot be identified to any of the known species	Tools for data discovery	Tools for identification	Tools for identification		
22	1	Researcher, Scientist	generate an artificial intelligence (AI) algorithm based on a selection of specimens	can run the algorithm on images of other institutes to verify identifications and recover specimens under my species concept	an AI toolbox to generate the AI algorithm	Tools for data analysis	Tools for identification	Tools for identification		
23	1	Citizen Scientist	identify a specimen	know which organism I have encountered	an automated identification tool	Tools for data analysis	Tools for identification	Tools for identification		
24	1	Researcher, Scientist	expose the collection I am contributing to other users, according to the FAIR principles	can collaborate with other users on new science using the collection	an overview of existing collection(s) with standard descriptions	Tools for data discovery	Metadata on collection level			
25	1	Researcher, Scientist	find specimens that have DNA data	can analyze past genetic diversity from subfossil remains	ancient DNA from collection specimens	Ecological data	Molecular data			

26	1	Data Scientist	explore the use of other data types provided by a standard, that are currently not used by a specific community	collections can be described using data types that were not explored before	any data type that was not used by a specific comunity, but is used by others	Tools for data discovery				
27	1	Site Manager CS	select a load of images	to build a CS project	basic elements on the images	Images				
28	1	Researcher, Scientist	gather / compare / reuse data from individual "small scale" (e.g. single species) studies	can do meta-analyses and find "large scale" general patterns	bulk data in comparable format; selective search & export tool; links to other databases (e.g. genbank, vegetation databases)	Legal and policy framework				
29	1	Researcher, Scientist	look at the cause of the current extinction catastrophe	can recommend action against it	clear and recent (within 20 years) monitoring data of biodiversity	Ecological data	Distribution data			
30	1	Researcher, Scientist	find out if a collection holds specimens of interest	can study the specimens	collection level data to make estimations about holdings data	Metadata on collection level				
31	1	Researcher, Scientist	find all physical specimens relevant to my field of interest, regardless of their digitization status	can examine the specimens for my research, either by visiting, requesting a loan, or possibly asking for imaging / digitization where possible	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, possibly list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata	Morphological data	Physical access		
32	1	Researcher, Scientist	compile species collection localities	can examine species distribution	collection site text string and geographic coordinates	Tools for data analysis	Distribution data	Tools for geo-referencing		
33	1	Researcher, Scientist	know where species occur based on museum collections	create predictions of their future distibution under global warming	collections digitized, location, date	Distribution data	Ecological data	Morphological data		
34	1	Researcher, Scientist	find out all specimens from a certain locality and date	make sure who collected it	collections digitized, location, date	Distribution data	Ecological data	Morphological data		
35	1	Researcher, Scientist	search for trait information of a certain species	study the changes in its traits based on changes in the environment due to global warming	collections digitized, location, date, high-resolution images of specimens, trait information	Distribution data	Ecological data	Morphological data	2D images	
36	1	Researcher, Scientist	know where a species occurs based on museum collections and occurences	can study what factors control its distribution range and the changes in its distribution due to habitat loss, also studu its extinction risk	collections digitized, observations, location, date, abiotic factors linked to specimens	Distribution data	Ecological data	Morphological data		
37	1	Historian The history of natural history	trace the life and daily practices of the person through collection data	write the biography of a natural historian/collector.	collector, photos (scientific)	Images related to collections	Metadata			
38	1	Researcher, Scientist	contact data of individuals of an herbarium	they can be contacted personally	contact data of individual persons connected to an herbarium	Metadata on collection level				
39	1	Citizen Scientist	be recognized as contributor	can apply for funding to digitize my own collections and can identify my contribution on validating data from external sources	contribution indicators (as contributor/validator)	Metadata on record level				

40	1	Researcher, Scientist	be recognized as contributor	can apply for funding to digitize any collection of special interest for research	contribution indicators (as contributor/validator)	Metadata on record level				
41	1	Researcher, Scientist	know the state of preservation	get genomic data from a specimen	data on the preservation methodology and sample age	Metadata on record level				
42	1	Researcher, Scientist	know legal procedure and limits for loan of preserved organisms and transport limits regarding dangerous goods for countries that host	choose material for scientific work also by simplicity of loan procedure	database of legal and transport procedures and limits for loan of preserved organisms by countries	Legal and policy framework				
43	1	Researcher, Scientist	extract data on presence and storage of all collected specimens of species of my present interest and detailes of loan process at insitutions that host the material	can easily choose and loan material for scientific work	database on presence and storage of specimens of species and host institution information on loan process	Legal and policy framework				
44	1	Researcher, Scientist	to model South East Asian biodiversity patterns	can predict biodiversity values and climate change impacts for entire regions including undercollected areas	detailed taxonomic and geographic information.	Distribution data	Morphological data	Tools for data analysis		
45	1	Researcher, Scientist	Model South East Asian biodiversity patterns		Detailed taxonomic and geographic information.	Distribution data	Morphological data	Tools for data analysis		
46	1	Researcher, Scientist	generate distribution maps	can check niche parameter	detailled taxonomic and geographic information	Distribution data	Ecological data	Morphological data		
47	1	Researcher, Scientist	revise a genus or part of a genus	can write a monograph	digital image of all specimens belonging to a genus or to a subset of a genus (section, geographical area) with tabulated information (extracted from the label and eventualy automaticaly exctracted from the specimen)	2D images				
48	1	Researcher, Scientist	check the identity of the specimen	can confirm the identification	digital images of species	2D images	Tools for identification			
49	1	Researcher, Scientist	check that the transcribed label data corresponds to the actual label information	can confirm collection details	digital images of specimen labels	Label images				
50	1	Researcher, Scientist	check if specimens have particular organs so that you can decide if it is valuable to look at the actual specimen	can visit institutes that have relevant specimens	digital images of specimens of certain species	2D images				
51	1	Paleobiologist	identify key specimens for paleobiological comparison	reassess the taxonomy of fossil mammals from India, the biogeography of South Asian mammals, and paleocommunity change	digitized catalogues of fossil mammals, reptiles, amphibians and birds for Henry Woodward, the first formal descriptive documents on Falconer’s Siwalik fossils including lectotypes	Tools for data analysis	Images related to collections	Tools for identification		

52	1	Entomologist	access original species descriptions and original type material, particularly older taxonomic works	compare written descriptions to the material I have access to for taxonomic revision	digitized custom-downloadable taxonomic descriptions allows faster and easier access to papers and allows me to quickly locate original descriptions from early taxonomic authors, particularly in old, rare, or large books	Data integration	Tools for downloading data/metadata			
53	1	Researcher, Scientist	see where and when species have occurred	create a forecast model about biodiversity changes	digitized specimen data	Ecological data	Distribution data	Tools for analysis		
54	1	Archaeologist	determine species of ivory tusk (African or Asian)	determine population migration and trading routes	DNA Fingerprint / barcoding	Molecular data				
55	1	Researcher, Scientist	know the DNA sequence of a specific specimen	study its phylogenomy	DNA sequence data	Molecular data				
56	1	Anthropologist, Archaeologist, Paleontologist	find ancient DNA analysis to identify genetic characteristics	determine origin and migration routes of contemporary aboriginal populations.	DNA. Osteologic collections as an anthropologic archive.	Molecular data				
57	1	Researcher, Scientist	learn how to correctly format a specimen Institution and Collection Code	can include references to a specimen in a scientific publication and properly cite voucher specimens in a GenBank submission	dwc:institutionCode dwc:collectionCode	Metadata on collection level				
58	1	Researcher, Scientist	find out if a specimen from a collection that no longer exists has been absorbed by another biorepository	can access the specimen or other data associated with it	dwc:institutionCode dwc:collectionCode institution address institution contact/s information about collections being / been transferred	Metadata on collection level				
59	1	Researcher, Scientist	find basic information about a particular collection in a particular biorepository or basic information about a particular biorepository	can discover more about it via a webpage OR via a webservice	dwc:institutionCode dwc:collectionCode institution address institution contact/s institution domain name institution website/s webservices (if they exist) information about collections being / been transferred metadata about the institution metadata about collection/s held	Metadata on collection level				
60	1	Curious person	know more about a specimen of interest	can further understand the other information related to this specimen in the place where I learned of said specimen (GBIF, publication, GenBank, website, etc)	dwc:institutionCode dwc:collectionCode metadata about the institution metadata about collection/s held	Metadata on collection level				
61	1	Citizen scientist	know where was a certain collector on a certain day	to help transcribe a specimen	existing transcription of specimen collected around the same time by the same	Metadata on record level				

62	1	Researcher, Scientist	find whether images available at institutional repositories are of taxonomic-grade	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	Metadata on record level	2D images			
63	1	Researcher, Scientist	find and reuse digital specimens from DiSSCo	use all digital specimens of a selection of species from DiSSCo for scientific research	fast access to the DiSSCo infrastructure	Tools for data discovery				
64	1	Invasive Species Controller (Environment administrator)	know where specimens of a certain invasive (plant) species occur	study why it has become so invasive / invasion history of the species	genetic data, collection data, location, date	Metadata on record level	Distribution data	Ecological data	Molecular data	
65	1	Public health official	know where bat/rat species occur based on collections and observations	identify and predict potential spread of a zoonotic disease (ebola, lassa fever) carried by those bats/rats	geographic range of collected specimens, location, date	Metadata on record level	Distribution data	Ecological data		
66	1	Researcher, Scientist	obtain specimen geo- and other data	can perform an IUCN Red list assessment	georeferenced specimens and label data on altitude, ecology, date	Metadata on record level	Distribution data	Ecological data		
67	1	Researcher, Scientist	find herbaria in foreign countries	collected voucher specimens can be deposited in the correct herbariim	herbarium location and expertise	Metadata on collection level				
68	1	Researcher, Scientist	check if a species I am interested in is new to science	can describe it as new to science	high quality images of the holotypes for the genus or group I wish to study, preferably with different relevant views (dorsal, ventral, detail shots, preparations and/or dissections if available)	2D images				
69	1	Researcher, Scientist	verify the validity of a determination by inspecting the holotype and/or other types for that species	can confirm the valid identity of a species	high quality images with different views (dorsal, ventral, detail shots etc...) and the metadata for the type specimens	2D images				
70	1	Association	to gather information to have overall figures representative of partners' state-of-the-art	can showcase the relevancy of the collections held to policy makers and attract funds	high-level figures that feature the collections as a whole	Legal and policy framework	Tools for reporting & statistics			
71	1	Researcher, Scientist	find out if a taxon has vanished from part of its geographic range / ecological niche	can think of causes, consequences, countermeasures	high-res image / DNA to verify species identification; collection metadata (year, location, taxon interactions, co-occuring taxa)	Tools for data discovery	Metadata			
72	1	Researcher, Scientist	find an image of a species to put in a press release	can allow journalists to write an article on my paper	images of living plants available with COO licence	Metadata on record level				
73	1	Citizen Scientist	help on transcribing and contribute	can enjoy	images without transcription	Metadata on record level				
74	1	Phytopatologi st	find pictures of verified digitized plant pathogens and pest symptoms on plants	create a solid basis for deep machine learning tool for automated pathogen detection	images, digitized specimen data	2D images				
75	1	Researcher, Scientist	explore the traits of a specific plant species	study the taxonomy of the species	images, sequence data, georeferences, traits	Metadata on record level				

76	1	Citizen Scientist, Site Manager	select multiple images	can build a CS project	information on the basic elements of the images	Metadata	Images			
77	1	Researcher, Scientist	focus on the native vs. not native status	can inform CS or other stakeholders	information on the status/biogeography	Tools for data discovery				
78	1	Researcher, Scientist	contact data of herbaria in specific countries	partnerships can be developed	institutional metadata	Metadata on collection level				
79	1	Anthropologist, Archaeologist, Paleontologist	find bone tissue with analyzed radioactive isotopes	attribute an age to finds	isotopes. Osteologic collections as an anthropologic archive.	Isotopic data				
80	1	Anthropologist, Archaeologist, Paleontologist	find analysis of stable isotopes of carbon and nitrogen on human skeletal remains	reconstruct diet of past populations	isotopes. Osteologic collections as an anthropologic archive.	Isotopic data				
81	1	Researcher, Scientist	know the quality of the data provides	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				
82	1	Researcher, Scientist	Map global species richness patterns	can model the correlates of species richness, determine drivers and integrate it with phylogenetic evidence	large numbers of accurately determined, georeferenced occurrences, with names standardised against globally accepted names standards	Metadata on record level				
83	1	Researcher, Scientist	look for specimens in different institutions collected from the same provenance	study colonialism / cultural history	links between collections from the same origin	Data integration				
84	1	Researcher, Scientist	look for links between living fungi collections and their digitized specimens	study the relationship between genetic properties and phenotypic traits	links between living specimens and digitized specimens	Data integration				
85	1	Researcher, Scientist	find all literature citing a specimen	can recontruct the taxonomic history	links between specimens and literature	Tools for data discovery	Data integration			
86	1	Researcher, Scientist	get an overview of previous knowledge about occurrence / genetic diversity of a taxon	can monitor genetic diversity of taxa across time & space with low resources (e.g. well-directed sampling of individuals, re-use of genetic marker systems)	links to DNA sequence data and publications, collection metadata (sampling location, taxon interactions), high-res image/DNA to verify species identification, access to original labels & additional sampling documentation (field notebook)	Tools for data discovery	Metadata on collection level			
87	1	Historian/researcher	find out all specimens and objects collected by a certain collector/explorer	study biography of the collector	list of all digitized specimens collected by a certain collector, links between the collections by the same collector	Tools for data discovery	Metadata on collection level			
88	1	Historian/researcher	see where and when a collector was collecting specimens	can write a biography on the collector	list of digitized specimens gathered by the collector: location, country, date of collecting, name of the collector, specimen ownership	Tools for data discovery	Metadata on collection level			

89	1	Researcher, Scientist	know in which botanic garden a specific plant can be found as a living collection	study its chemical traits	list of living plants in collections	Tools for data discovery				
90	1	Researcher, Scientist	know in which seed bank seeds of a specific plants are stored	study its germination	list of species stored as seeds in seed bank collections	Tools for data discovery				
91	1	Researcher, Scientist	get a data dump on all institutions and their collections that can be sorted by		Location (country, city, state), Institutional name, Domain name, Type of institution (public, private non-profit, private for-profit, museum, herbarium, etc.), Type of collection preservation (dry/room temperature, liquid/room temperature, cryogenic), size of collection	Metadata on collection level				
92	1	Food Security Official	find herbarium specimens of a plant disease.	compare genomes to modern day disease outbreaks.	location of collections, genomic data (if existing), date.	Metadata on collection level				
93	1	Public health official	compare sequences of anthrax	identify the strain used in terror caused by sending anthrax letters	location of museum specimens to be used for genetic studies / sequences of those specimens	Tools for data discovery	Metadata on collection level	Data integration		
94	1	Researcher, Scientist, Public Health Official	locate museum specimens (eg. Egg shells) for study taxa	study the effects of environmental contaminants	location of museum specimens, number of specimens, location, date	Tools for data discovery	Metadata on collection level	Data integration		
95	1	Researcher, Scientist	look for basic information about particular collection within an institution / a particular museum/herbarium/other biorepository institution		Location: country, state, city; Institutional name, Institutional acronym, domain name, a webpage I can browse / webservice response	Metadata on collection level				
96	1	Researcher, Scientist	compare the community compositions in water samples based on metabarcoding data	can assess the impact of abiotic and anthropogenic drivers on community compositions	metabarcoding sequence data	Interoperability	Molecular data			
97	1	Citizen Scientist	find specific info	can use it in blogs, publications etc.	metadata and photos	Metadata	Advanced search functionality			
98	1	Researcher, Scientist	do biographical research on specific collectors	contribute to cultural history (e.g. colonial history)	names of collectors and places of origin of specimens	Metadata on collection level	Data integration			
99	1	Researcher, Scientist	explore where specimen were found and at which time they lived there	analyse spatio-temporal turnover in relation to environmental drivers	nice visualisation tools of all specimen, which can be filtered by organism group or specimen	Tools for data discovery	Tools for data visualization			
100	1	Anthropologis t	understand object's temporal context	evaluate changes in the making of objects	object measurements, historic owners	Tools for data analysis	Metadata on record level			
101	1	Researcher, Scientist	know the newest occurences of a certain alien species	provide an impact assesment for decision makers	occurence data, date, location, observations verified, number of individuals, timing & frequency of observations, uncertainty, distribution of a	Tools for data analysis	Distribution data			

102	1	Historian/researcher	know where plants from the genus Solanum are being cultivated based on botanical collection data	can study the impact of introduction of Solanum plants to traditional agriculture	occurrence data of Solanum plants within countries where they have been introduced, list of introduced Solanum species per country	Tools for data analysis	Distribution data			
103	1	Researcher, Scientist	get the data for Agrimonia eupatoria occurences	create a species distribution model	occurrence data of the plant: location, date, other metadata linked to observations/specimens, observations to be verified	Tools for data analysis	Distribution data			
104	1	Researcher, Scientist	see the where species from a certain genus occur	can create a distribution map of each species	occurrence data, location, date	Tools for data analysis	Distribution data			
105	1	Researcher, Scientist, Public Health Official	know where a mosquito species occurs / where it has been collected	predict risk of new infections based on its distribution	occurrence data, location, date	Tools for data analysis	Distribution data			
106	1	Researcher, Scientist	know all species occurring in an area and their population status	create a red-list	occurrence data, location, date, number of individuals, timing and frequency of observations, uncertainty	Tools for data analysis	Distribution data			
107	1	Anthropologist, Archaeologist, Paleontologist	use digitized information on skeletal lesions	investigate the type of work done and formulate hypotheses about the economy, lifestyle, and division of labor within ancient societies	osteologic collections as an anthropologic archive	Tools for data analysis	Morphological data			
108	1	Anthropologist, Archaeologist, Paleontologist	find macroscopic, histologic and radiographic studies of lesions on a skeleton.	trace diseases affecting past populations and understand their origin and diffusion through identification of bacterial DNA	osteologic collections as an anthropologic archive	Tools for data analysis	Morphological data			
109	1	Anthropologist, Archaeologist, Paleontologist	find genetically determined characteristics (structure of dental tubercles and non-metric skeletal)	study origin of populations and relations between them	osteologic collections as an anthropologic archive	Tools for data analysis	Morphological data			
110	1	Forensic Facial Reconstructionist Facial 3D Modeling		can bring a face from the past alive (famous historical figures, archaeological dig). Determine if two skulls from same family or population	osteologic collections as an anthropologic archive, DNA, and photos.	Tools for data analysis	Morphological data			
111	1	Anthropologist, Archaeologist, Paleontologist Human Evolution	access their catalogues, digital pictures and drawings (for old collections) as well as any other useful information for my research such as diaries, notes, and letters	use fossil bones and lithic collections in museum and university collections upon which my research relies	part description, place of collection or georeferenced locality, date of collection, object measurements, photos (scientific)	Tools for data discovery	Metadata on record level			

112	1	Historian Geospatial Science		understand historical disseminations of plants and animals, and long-term human influence on these; as well as how these collections have been constructed in a colonial/imperial/globalization context	part description, place of collection or georeferenced locality, date of collection, collector	Tools for data discovery	Metadata on record level			
113	1	Art Historian	identify organic materials used in art (wood frames, paints)	identify species and age of art materials provides temporal, social context	reference collection, expertise, or DNA fingerprint	Metadata on record level	Molecular data	Tools for identification		
114	1	Paleontologist, Earth & Life Scientist	access well preserved representative collection specimens to score morphological characters to take samples for genetic and genomic research	better understand the evolution of organisms	scientific name, part description, place of collection or georeferenced locality, collector, DNA, tissue sample	Tools for data discovery	Molecular data	Morphological data	Distribution data	
115	1	Paleontologist, Earth & Life Scientist	access to type materials	link taxon names to clades	scientific name, part description, place of collection or georeferenced locality, collector, DNA, tissue sample	Tools for data discovery	Molecular data	Morphological data	Distribution data	
116	1	Anthropologist, Ethnobotanist Medicinal and ritual plants	have access to field notes or accession books as the information is often not written on object labels or not digitized from labels.	compare plant uses	scientific name, part description, place of collection or georeferenced locality, date of collection, collector, photos (scientific), historic owners, DNA	Tools for data discovery	Molecular data	Morphological data	Distribution data	Images related to collections
117	1	Anthropologist, Ethnobotanist Medicinal and ritual plants	have access to associated data which are often not mentioned on the labels or digitized from the labels.	study historic rice specimens	scientific name, part description, place of collection or georeferenced locality, date of collection, collector, photos (scientific), historic owners, DNA	Tools for data discovery	Molecular data	Morphological data	Distribution data	Images related to collections
118	1	Anthropologist, Archaeologist, Historian, Linguist / Languages Spatial Humanities	track the collection, use and dissemination of the resource, and understand its cultural value.	investigate the historical exploitation of resources	scientific name, part description, place of collection or georeferenced locality, date of collection, object measurements	Tools for data discovery	Morphological data	Distribution data		

119	1	Anthropologist, Archaeologist Long-term human-environment, especially human-animal interaction, during pre-Columbian and early Historic Era times	quantify the species' presence across sites, create chronological context, and conduct morphological comparisons across individuals	understand the human impact on the spatial, temporal and cultural distribution and use of a species	scientific name, part description, place of collection or georeferenced locality, date of collection, object measurements, DNA, isotopes, absolute or relative date of specimen life (not age or date of collections)	Tools for data discovery	Molecular data	Morphological data	Distribution data	
120	1	Botanist Cross-cultural Ethnobotany	verify and comprehensively compare my data with authentic online easily available resources		scientific name, part description, place of collection or georeferenced locality, preparation, photos (scientific)	Metadata on record level	Morphological data	Distribution data	2D images	
121	1	Botanist Cross-cultural Ethnobotany	have access to a comprehensive stock of online resources	collect ethnobotanical data and specimen, and be able to compare the data with previous literature	scientific name, part description, place of collection or georeferenced locality, preparation, photos (scientific)	Metadata on record level	Morphological data	Distribution data	2D images	Images related to collections
122	1	Paleontologist Evolution and Systematics of Ungulates	know the species present in a locality (for zooarchaeological studies), the age of the site, and the environment	study the evolution of animals in Prehistory	scientific name, place of collection or georeferenced locality, (I) weight and body length of live specimen. Information available in a public catalogue. The objects label in the collection, should help identify the specimen and protect it against being mislaid.	Metadata on record level	Morphological data	Distribution data		
123	1	Paleontologist, Geologist Taphonomy	identify substrate level preferences of genera	analyze the impact of climate change on generic diversity, paleocommunities, and substrate affinity	scientific name, place of collection or georeferenced locality, collector, object measurements, photos (scientific), stratigraphic and geological age attributes	Metadata on record level	Morphological data	Distribution data		
124	1	Anthropologist Ethnoecology	use reference collection objects, online resources and associated archives (photos) to identify a species	determine the natural species used in the construction of objects	scientific name, place of collection or georeferenced locality, date of collection, collector, photos (scientific)	Metadata on record level	Morphological data	Distribution data	2D images	Tools for data discovery
125	1	Anthropologist Ethnoecology	have physical or online access to similar collections including data and photos	compare objects between our collections and similar collections in other	scientific name, place of collection or georeferenced locality, date of collection, collector, photos (scientific)	Metadata on record level	Morphological data	Distribution data	2D images	Tools for data discovery
126	1	Anthropologist Ethnoecology	check the existence, date of sampling, local names, description of uses, etc. of date palms in collections	confirm presence and use of plants (date palms for instance)	scientific name, place of collection or georeferenced locality, date of collection, collector, photos (scientific), DNA	Metadata on record level	Morphological data	Distribution data	Molecular data	Tools for data discovery

127	1	Historian Cultural Heritage	look up archival resources using data on the object label	establish the history of science (taxonomy in the 19th century)	scientific name, place of collection or georeferenced locality, date of collection, collector, preparation	Metadata on record level	Morphological data	Distribution data	Tools for data discovery	
128	1	Historian	locate all specimens and objects associated with a particular expedition or exhibition.	investigate the collection of a particular species of crocodile from an expedition to Borneo around 1841	scientific name, place of collection or georeferenced locality, date of collection, collector, preparation, photos (scientific), historic owners, cause of death	Metadata on record level	Morphological data	Distribution data	Tools for data discovery	
129	1	Historical Ecologist / Environmental Historian	obtain occurrence and abundance data in the past	reconstruct biodiversity in the past	scientific name, place of collection or georeferenced locality, date of collection, object measurements, cause of death	Metadata on record level	Morphological data	Distribution data	Tools for data discovery	
130	1	Historian Cultural Heritage	access archives using data on the object label	trace the origin, movement and history of collections	scientific name, place of collection or georeferenced locality, date of collection,collector, historic owners	Metadata on record level	Morphological data	Distribution data	Tools for data discovery	
131	1	Paleontologist, Earth & Life Scientist	access collection catalogues, accession books, literature, or where provenance is unclear, biographic material, field notes, and historical maps	understand what previous researchers meant by a specific taxon name	scientific name, place of collection or georeferenced locality, date of collection,collector, historic owners	Metadata on record level	Morphological data	Distribution data	Tools for data discovery	Images related to collections
132	1	Anthropologist Ethnoecology	check all mentions of plants (with date, attribution) used, collected, traded, etc. in a given area	analyze the evolution of agriculture in an oasis	scientific name, place of collection or georeferenced locality, date of collection,collector, photos (scientific), historic owners	Metadata on record level	Morphological data	Distribution data	2D images	Tools for data discovery
133	1	Anthropologist, Historian History of Science	know exactly where and when collected, what kind of data (e.g., vernacular names, uses) was collected, and subsequent movements of the specimens including public display in galleries where this occurred	understand how a group of ethnobotanical specimens was collected and what this tells us about (a) the source community and (b) the motivations of the collectors and subsequent institutions	scientific name, place of collection or georeferenced locality, date of collection,collector, photos (scientific), historic owners	Metadata on record level	Morphological data	Distribution data	2D images	Tools for data discovery
134	1	Anthropologist, Historian History of Science	check historic plant resources to enable me to determine what was available in the region and what it looks like.	check identification of plant material in an ethnographic object	scientific name, place of collection or georeferenced locality, date of collection,collector, photos (scientific), historic owners	Metadata on record level	Tools for identification	Morphological data	Distribution data	2D images
135	1	Historian The history of natural history	ascertain all relevant historical information on an object	write an object biography of an important object from a natural history museum	scientific name, place of collection or georeferenced locality, date of collection,collector, photos (scientific), historic owners	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	2D images
136	1	Historian	locate digitized specimen collection records, associated photos and archives describing the collecting event.	investigate the collection of a particular species of crocodile from an expedition to Borneo around 1840	scientific name, place of collection or georeferenced locality, date of collection,collector, photos (scientific),historic owners. Other: societal context of object at the time of collection.	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections

137	1	Geologist		fill-in data gaps regarding specimens and collections with limited (species name and locality) data	scientific name, place of collection or georeferenced locality	Metadata on record level	Distribution data	Tools for data discovery		
138	1	Anthropologist, Historian, Philosopher, Sociologist Science fiction, Datafication of nature		follow the epistemic life of an object as it is involved and shapes research practices	scientific name, sex / age part description place of collection or georeferenced locality, date of collection, photos (scientific), DNA, geological references	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Molecular data
139	1	Anthropologist Physical Anthropology	check the accession books, field notes or diaries, and possibly correspondence and notes	study the history of collections	scientific name, sex / age part description, place of collection or georeferenced, locality, collector cause of death	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections
140	1	Anthropologist, Archaeologist Long-term human-environment, especially human-animal interaction, during pre-Columbian and early Historic Era times	quantify target taxa across a site(s), determine age and sex to create mortality and demographic profiles of the taxa, conduct aDNA to assess impacts of human influence on population genetic diversity, and use isotopic analysis to assess diet.	identify patterns of animal management or incipient domestication	scientific name, sex / age, DNA, isotopes	Metadata on record level	Tools for data discovery	Isotopic data	Distribution data	Molecular data
141	1	Anthropologist Physical Anthropology	Identify the field notes and drawings of the archaeologist	reconstruct a prehistorical population	scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, historic owners, cause of death	Tools for data discovery	Metadata on record level	Morphological data	Distribution data	Images related to collections
142	1	Anthropologist, Archaeologist, Paleontologist Human Evolution	compile information such as scientific name, dating, geographic information on the localities where they were found, basic quantitative and qualitative description. Field notes, illustration (drawings and photos), and any other useful information about them that was published or unpublished.	research fossils and prehistoric stone tools	scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, object measurements, photos (scientific),	Tools for data discovery	Metadata on record level	Morphological data	Distribution data	Images related to collections

143	1	Ethnobotanist Biocultural Heritage	use information from historical collections that have information to be integrated.	infer about the history of biodiversity in the area of the Brazilian Amazon	scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, preparation, object measurements, photos (scientific)	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections
144	1	Classicist, Linguist, Languages Literature, Digital Philology	make augmented text searches; establish georeferenced localities	connect realia to encyclopedic information for historical documents (e.g. WWI & WWII documents or ancient Greek documents)	scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, preparation, object measurements, photos (scientific), historic owners	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections
145	1	Anthropologist, Historian, Philosopher, Sociologist Science fiction, Datafication of nature	understand the transforming objectivities [sic] of specimen in natural history collections, and trace them across different media and databases		scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, preparation, object measurements, photos (scientific), historic owners, cause of death	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections
146	1	Historian History of Sciences, Scientific Illustrations, History of Photography	use all available resources, for example, a specimen's sex, age and date of collection is relevant for identifying a possible bias of the collectors/collection during a certain period of time	put a collection in socio-historical context	scientific name, sex / age, part description, place of collection or georeferenced locality, date of collection, collector, preparation, object measurements, photos (scientific), historic owners, cause of death	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	Images related to collections
147	1	Historian History of Sciences, Scientific Illustrations, History of Photography	Use all available sources of information, images, text (publications, correspondence, notes, etc.) audio, or moving image of an object / subject	fully understand/trace an object or subject's history (how an object came to the institution, provenance, circumstances of collecting event)	scientific name, sex / age, part description, place of collection or georeferenced, locality, date of collection, collector, preparation, object measurements, photos (scientific), historic owners, cause of death, provenance form of acquisition	Metadata on record level	Data integration	Morphological data	Distribution data	Images related to collections
148	1	Sociologist Cultural policy			scientific name, sex / age, place of collection or georeferenced locality, date of collection	Metadata on record level	Tools for data discovery	Morphological data	Distribution data	
149	1	Researcher, Scientist	find specimens matching a DNA sequence I have generated	can identify collected material similar to my sample (and thereby identify my sample)	searchable records of DNA sequence linked to collected material	Tools for data discovery	Molecular data			
150	1	Public health official	look for viruses found in animal tissues in collections	identify an unknown vector for a disease	sequence data, links between collections	Metadata on record level	Data integration	Molecular data		

151	1	Paleontologist, Geologist Taphonomy	easily discover specimen-related literature and speed data recovery		single database interface linking specimens with their literature or metadata	Metadata	Data integration			
152	1	Paleontologist, Geologist Taphonomy	access georeferenced modern and fossil occurrence data for species throughout their geographic distributions in the Atlantic Verify taxonomy of each species occurrence and pair with temperature and salinity measurements	study how climatic changes drove thermophilous mollusk species migrations during the Pleistocene	single database interface linking specimens with their literature or metadata	Metadata on record level	Data integration			
153	1	Researcher, Scientist	previouly collected specimens with their taxon and geographical area	the relevance of additional studies can be determined	specimen taxa and location of collection	Tools for data discovery	Distribution data			
154	1	Data Scientist	create novel views of the data	get added value out of linking different types of information, e.g. collection data with phylogenetic trees, sequence data, literature citations, travel reports etc	stable, reliable identifiers of collections and collection objects (specimens)	Tools for data discovery	Data integration			
155	1	Researcher, Scientist	query when and where one or more species have been recorded, and their characteristics, and the institutions that archive specimens	can collect more specimens, or borrow collections	taxonomic fields, geographic coordinates, date	Tools for data analysis	Morphological data	Distribution data	Ecological data	Data integration
156	1	Researcher, Scientist	query when and where one or more species have been recorded, what their characteristics are, and which institutions archive the specimens	can use more specimens, or borrow collections from other institutes	taxonomic information, geographic coordinates, date of collecting	Tools for data analysis	Morphological data	Distribution data	Ecological data	Data integration
157	1	Food Security Official	see reference specimens for pests	can identify origins of newly found pest populations and study possible bioterrorism	temporal and geographic range of existing specimens	Metadata on record level				
158	1	Researcher, Scientist	use samples from zoological specimens in museums and zoos	can use this information for genetic management of ex-situ conservation breeding programmes and can develop science-based conservation action plans for endangered species	tissue samples, DNA, and/or sequence data, taxonomic information	Interoperability	Morphological data	Molecular data	Physical access	Data integration
159	1	Researcher, Scientist	have an overview of the geographic/spatial coverage of a certain taxonomic part of a collection	can assess the need for gathering further data or evaluate the usefulness of a certain collection for a certain analysis	to be able to filter specimens by collection, area and taxonomic identity	Metadata on collection level	Tools for data discovery			
160	1	Researcher, Scientist	to find type specimens	can verify and understand its taxonimic concept	to be sure this name is understood uniformerly	Tools for data discovery				
161	1	Researcher, Scientist	find which pathogens share geographic distributions with a taxon of interest	can filter potential pathogens that cause a disease	to combine different data sources of species occurrences	Tools for data analysis	Distribution data			
162	1	Researcher, Scientist	find which pollinators share geographic distributions with plants	can construct plant-pollinator networks	to combine different data sources of species occurrences	Tools for data analysis	Distribution data			

163	1	Researcher, Scientist	study the details of the organs	can identify the specimen as with a binocular	to download images with a resolution of 600 dpi minimum (300 dpi is insufficient).	2D images	Tools for downloading data/metadata	Tools for identification		
164	1	Researcher, Scientist	destructively sample for pollen, anatomy and DNA	can construct phylogenies and understand evolution	to know policies associated with each specimen.	Legal and policy				
165	1	Researcher, Scientist	know where the specimens are kept in the collection	can find them in the collection	to know the number of the boxes where the specimens are kept	Metadata on record level				
166	1	Researcher, Scientist	get data from specimens of species from which I obtained trait data through EOL Traitbank	can study geographical patterns of (functional) traits	to query DiSSCo through EOL traitbank selections	Tools for data discovery	Interoperability	Ecological data		
167	1	Researcher, Scientist	get data from specimens from which I obtained sequence data through genbank	can study geographical patterns of genetic variation	to query DiSSCo through Genbank accessions	Tools for data discovery	Interoperability	Molecular data		
168	1	Researcher, Scientist	extract data on the variation in specimens of the same species	can study trait variation within a species	to select trait data from specimens	Tools for data discovery	Ecological data			
169	1	Researcher, Scientist	extract species occurrence data in a particular location or area	see whether data exist in the first place and if exist use it for analyses of spatial/temporal variation in biodiversity	tool to define the extracting and data retrieval system	Tools for data analysis	Distribution data			
170	1	Researcher, Scientist	extract trait data, which might include phenology and morphology	can analyse the phenology or trait variation within species	trait data of species, either from the specimens or from EOL traitbank	Tools for data discovery	Interoperability	Ecological data	Morphological data	Tools for data analysis
171	1	Researcher, Scientist	search for all type specimens of a certain genus	verify that a species I think is new to science is really that	type specimens digitized with high-resolution images, annotations, synonyms, species characters, genetic data, phytochemistry	Tools for data discovery	Label images	2D images	Morphological data	Molecular data
172	2	Collection Manager		make a locality check or search for old and correct names for labeling	(historic) maps	Metadata	Images related to collections	Tools for data discovery		
173	2	Digitization Officer	link specimen & label images the corresponding occurence data	can make the images as publicly visible and usable as possible	a (semi)automatic label data extraction & verification system	Tools for downloading data/metadata	Distribution data	Label images	2D images	Data integration
174	2	Collection Manager	receive information on the GeneBank numbers and analysed molecular data	can link this back to specimens, tissue or DNA samples	a connection of genetic data (e.g. shared genetic libraries) which are interlinked with specimen database entries at my home institution	Advanced search functionality	Data integration	Interoperability		
175	2	Curator	attach and deliver geochemical data to rock and mineral specimen records	retrieve specimens based on their geochemical signature	a portal compatible or similar to the Earth Chem portal funded by NSF http://www.earthchem.org/	Advanced search functionality	Tools for uploading	Biochemical or geochemical data	Data integration	
176	2	Curator	monitor and in specific cases restrict access to geographical coordinates of collection sites	stop ruthless exploitation of strictly protected species	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				
177	2	Curator	enrich my collection with reliable annotation from specialists anywahere in the world	can increase the valua of my collection	a quality / reliability rating of annotators	Tools for annotation				

178	2	Curator	publish my data online	can increase the value of the collection	a user friendly collection management system (CMS)	Tools for uploading				
179	2	Collection Manager	be able to access information about storage conditions/status for specimens	can optimise the storage of my own collections and plan for future acquisitions	access to storage/status information of specimens	Tools for data discovery	Metadata on collection level			
180	2	Citizen Scientist	add label information to the specimen records	can contribute to scientific data	access to the DiSSCo portal	Tools for annotation	Tools for uploading			
181	2	Museum Preparator	know how old specimens are prepared in other collections	When restoring historical specimens we need as much information as possible about how they were prepared	accession books, collection catalogues, field notebooks / diaries, correspondence	Metadata on collection level	Images related to collections			
182	2	Collection Manager	access unique information such as orginal descriptions, pictures or species drawings	can ID (exotic) species in the collection	accession books, rare books / special collections, photos	Metadata on record level	Images related to collections			
183	2	Collection Manager	document the backlog (uncatalogued collections) of a cross-disciplinary not-for-profit museum.	to prioritize areas in need of digitization	accession records	Metadata on record level				
184	2	Digitization Officer	produce digital specimens from a digitisation line	can store a specimen in my collection management system (CMS) and can upload the images to a customer's CMS	an automated workflow minting persistent identifiers (PIDs)	Tools for uploading	Data integration			
185	2	Citizen Scientist	know where a certain collector was on a certain day	can help to transcribe a specimen	an existing transcription of a specimen that was collected around the same time by the same collector	Metadata on record level	Advanced search functionality			
186	2	Curator	annotate all aspects of records, suggest improvements, and record logical connections between records (link museum samples to corresponding occurence records)	can provide duplicate-free, reliable, well-documented data to end users	an record-level annotation and communication system spanning across institutes	Annotation tools	Data integration			
187	2	Collection Manager	have a tool to estimate undigitized backlog	hidden collections can be revealed, estimate the remaining effort for digitization	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, possibly list of taxon / species names, important collectors, thematic interest focus, digitization status	Tools for reporting & statistics	Metadata on collection level	Morphological data	Distribution data	Ecological data
188	2	Curator	answer multiple requests on a specified taxon / collector / geographic area	can follow conversation about a request	communication thread by taxon / collector / geographic area	Tools for clustering requests				
189	2	Curator	be recognised as contributor	can apply for funding to digitise institutional collections	contribution indicators (as contributor)	Tools for reporting & statistics				
190	2	Curator	attach relevant references to the specimen record	document curatorial decisions	custom downloadable references	Tools for uploading				

191	2	Curator	relate catalogue numbers of material in my collection to published scientific papers where they have	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	Advanced search functionality	Metadata on collection level	Data integration		
192	2	Collection Manager	check in which institutions certain collection categories are kept so that I can forward a collection on offer to an institute that is interested	can forward this information to a collection holder	details about taxonomic/geographic specialism and possibly wish lists	Advanced search functionality	Metadata on collection level			
193	2	Collection Manager	check in which institutions certain collection categories are kept	can forward a collection on offer to an institute that is interested	details about taxonomic/geographic specialism, and possibly wish lists	Advanced search functionality	Metadata on collection level			
194	2	Researcher, Scientist	correct an identification and add an annotation	can file the specimen under the correct taxonomic name	digital images and an annotation system	Annotation tools	2D images	Tools for uploading		
195	2	Citizen Scientist	read (handwritten) untranscribed label data	can add specimen details to the record	digital images of specimen labels	Label images	Annotation tools			
196	2	Curator	compare an unidentified specimen with identified specimens to determine their identity	identify the specimen	digital images of specimens	2D images	Tools for identification			
197	2	Curator	discover the type status of specimens	know how many type specimens are in the institutional collection	digitised information on the description of species	Metadata on record level	Tools for reporting & statistics			
198	2	Archive Manager		facilitate volunteer or crowd-sourced object registration and transcription	digitized accession books, field notebooks, collection catalogues	Images related to collections				
199	2	Museum Preparator	have quick reference from workstation, home computer, or phone.	reconstruct anatomically correct unique specimens	digitized detailed descriptions and illustrations of specimens	2D images	Metadata on record level			
200	2	Curator		find missing information on specimen labels	digitized literature	Label images	Images related to collections			
201	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	Data integration			
202	2	Manager	update or create institution or collection information	the metadata can be as accurate AND up-to-date as possible	dwc:institutionCode dwc:collectionCode institution address institution contact/s institution domain name institution website/s webservices (if they exist) information about collections being / been transferred metadata about the institution metadata about collection/s held	Metadata on collection level				
203	2	Historian	find information on the history of objects and collections	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.	Advanced search functionality	Tools for data discovery			

204	2	Citizen Scientist	help transcribing specimens	can enjoy	images of specimens without transcription	2D images	Annotation tools			
205	2	Director	hire a curator with knowledge of specific groups.	can be sure they have a background that includes knowledge of the main collection.	information on the importance of the collection gauged by size, taxonomy, scope, and time period.	Metadata on collection level				
206	2	Collection Manager	know which users are interested in which data	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...)	Metadata on collection level				
207	2	Collection Manager	create and/or update information on my institution and component collections		Institutional name and acronym, Collection name and code, Address and contact information, Domain name and websites, Webservices, Transfer of collections to new parent institutions, Characteristics of the institution and/or collection	Metadata on collection level				
208	2	Curator	add annotated information from an Unified Curation and Annotation System (UCAS) to my collection management system (CMS)	can update infomation on my specimens in my CMS	interoperability between my CMS and UCAS	Interoperability	Annotation tools			
209	2	Collection Manager	start a digitizing project	like to digitize a certain group of my collection; I like to do this internationally because of funding	know where else there are collections of this group	Advanced search functionality	Metadata on collection level	Tools for data discovery		
210	2	Collection Manager	forward a researcher to colleagues	they can exammine more collections	need to know which institute holds specific kinds of collection	Metadata on collection level				
211	2	Collection Manager	find original publication	identify type specimens	original literature	Metadata	Images related to collections	Tools for identification		
212	2	Citizen Scientist	keep track or my records and corresponding samples in a way that makes donating the specimens to a local museum painless to all parties	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.	Tools for documentation	Data integration			
213	2	Curator	use the taxonomic name finding feature to locate literature on specific species, download relevant documents as PDFs, and link to useful references using page permalinks	find type specimens in the collection, correctly identify them as types, and then update our records accordingly	quick access to old references and being able to download them	Advanced search functionality	Tools for downloading data/metadata			
214	2	Collection Manager	have access to primary data sources	can compare registration of data	rare books / special collections	Metadata	Tools for data discovery	Images related to collections		
215	2	Curator, Collection Manager	increase the collections visibility for the general public	can motivate lay people on the value of diveristy	specialized personnel to present parts of the collection in a story telling, yet scientifically sound, manner	Metadata on collection level				
216	2	Collection Manager, Administrator	use a unique identifier for a collection (e.g. ASIH codes for fish collections)	collections can be easily identified	stable identifier of a collection	Metadata on collection level				

217	2	Collection Manager	have the highest possible level of data security	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			
218	2	Curator	images (old or modern photos or drawings of the complete specimen or of any microscopic technique applied on it)	in order to make the accurate identification	these images are extracted from the original publications or requested from their authors and stored	2D images	Images related to collections		
219	2	Curator	annotate digital specimens with updated determinations	improve the curation of the collection	to be able to annotate a digital specimen and pass that to the curating institute	Annotation tools	2D images		
220	2	Director	know how much my institution's collections is used and for what	can argue for the importance of my institution's collection	to be able to extract information from DiSSCo based on views/downloads/annotations etc. of my collection.	Annotation tools	Tools for downloading data/metadata	Metadata on collection level	Tools for reporting & statistics	Legal and policy framework
221	2	Collection Manager	profile my collection	can show the importance of the collection	to be able to highlight my institutional collection within the DiSSCo collection	Metadata on collection level			
222	2	Collection Manager	analyse the uniqueness of the institutional collection	can communicate its value	to compare my collection with data on collections from other institutes	Tools for data analysis	Metadata on collection level	Tools for reporting & statistics	
223	2	Curator	find material for a given taxon	can pull material on taxonomic basis incl. outdated taxonomic view	to get a list of all possible floristic and taxonomic concepts involved	Advanced search functionality	Morphological data	Tools for data discovery	Tools for data analysis
224	2	Collection Manager, Director, Administrator	know the situation with collection amounts	can plan ahead for future storage needs	to know existing amounts of collections, and amount of new material coming in. Also, need to know status / condition (wet, dry, ...) of existing material. Also collection health information.	Advanced search functionality	Metadata on collection level	Tools for data discovery	
225	2	Collection Manager	understand the needs of researchers	make information available useful, and develop collections effectively	to know what researchers need	Metadata on collection level	Advanced search functionality		
226	2	Digitization Officer	ensure digitisation serves research needs	make effective use of resources	to know what researchers require	Advanced search functionality			
227	2	Collection Manager	connect a researcher to colleagues	they can examine more collections	to know which institute holds specific collections	Metadata on collection level			
228	2	Collection Manager	start a digitization project	can digitize a certain group of my collection; I like to do this internationally because of funding	to know which other institutes hold collections of this group	Metadata on collection level			
229	2	Collection Manager	check status of specimens collected in regard to Nagoya protocol	ask for collect authorization to scientists and check the law applied in the collect country	to link the digitized image to a scan of the collect authorization if needed	Legal and policy framework	Interoperability	2D images	Data integration
230	2	Collection Manager	encourage remote curation of my collection through expert annotation	can improve the curation and value of the collection	to receive and be able to easily incorporate annotated data	Annotation tools			

231	2	Curator	answer requests for specified objects	can search the collection and pull material	to receive requests including all taxon names involved	Tools for data discovery	Tools for reporting & statistics			
232	2	Researcher, Scientist	extract handwriting samples of a collector	verify collection localities and collection dates of specimens of a collector	to select all digitized labels from a specific collector	Tools for data discovery	Label images			
233	2	Curator	cross-check data between specimens collected by the same collector on the same day	can confirm that all specimens have similar geographical coordinates, or correct coordinates where necessary	to select all DiSSCo records by collector and date	Tools for data discovery	Tools for geo-referencing			
234	2	Researcher, Scientist	cross-check data between specimens of the same species	can flag outliers and correct record data where necessary	to select all DiSSCo records from a certain species	Tools for data discovery				
235	2	Collection Manager	measure the use of collections via citations	understand the use of the collection and give evidence of its importance	track specimen identifiers and their citation	Tools for data discovery	Tools for reporting & statistics			
236	3	Georeferencer	Refine or verify collecting locality	can be able to add georeferenced occurrence points to locality data	Digitized accession books, collection catalogues, and field notebooks	Metadata	Images related to collections	Tools for geo-referencing		
237	3	IT support	build and provide solutions and related services	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. Even better it would be if there is an institution digitisation priority list.	Metadata on collection level	Tools for reporting & statistics	Tools for data discovery		
238	3	Data service provider	check and normalize the often diffuse text references that turn up in specimen data aggregates against well-managed vocabularies	can rely on the standard collection codes and other identifiers derived from the catalogue to link to other, external sources of relevant data (e.g. literature, sequences, traits, etc)	institutionCode, collectionCode, other collection identifiers; country / isoCountryCode, named polygons and base of reference where available, field collection protocol (standardized catalog), taxonomic scope, temporal scope, collector names	Metadata on collection level	Interoperability	Data integration		
239	3	Biodiversity Data Node Manager	a platform for the aggregation and retrieval of Biodiversity data (both collection and observation data)	provide access to Biodiversity data and data analysis tools	interaction with various biological data publishing institutions and organisations,	Interoperability	Tools for data discovery	Tools for data analysis	Data integration	
240	3	Solution provider	tap into the vast market of digital storage solutions for digital natural collections	can sell my services and consult	predictable numbers on collection type, volume and progress in digitization	Metadata on collection level	Tools for data discovery	Tools for reporting & statistics		
241	3	Software developer	ceate new usages with the data and ways to add to the data, through apps or web interaction	the data is easier accesible 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.	scope: collection level - details: specimen level	Metadata on collection level	Interoperability	Advanced search functionality	Data integration	

242	4	Policy maker	to understand the value of DiSSCo	can justify the national level investments in its operations	access to impact stories and/or assessments	Tools for data discovery	Legal and policy			
243	4	Policy maker	Know how well sampled my country is	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.	Tools for data discovery	Tools for data visualization	Tools for geo-referencing		
244	4	Policy maker, Funder [GBIF Node Manager?]	prioritize collections for digitization and mobilization, demonstrate the importance of particular collections to fill known data gaps	funding is mobilized for digitization and invested at the point of best value to broaden the base for further decision making; institutions prioritize digitization in their work plan, infrastructure for digitization is developed	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data discovery	Tools for reporting & statistics		
245	4	Administrator	compare interdisplinary collections	am able to compare two collections from different communities	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Data integration		
246	4	Collection Manager	use data about biodiversity, conservation; behavior and life history	the data can be used to help soliciting support	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, possibly list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis			
247	4	Collection Manager	have an easy tool to visualise the data of my collection	data can be presented easily to other stakeholders (e.g. funding agencies)	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, possibly list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Tools for data visualization		
248	4	Collection Manager, Administrator	compare different collections. Find unique characteristics of a collection	the data can be used as a tool to help secure funding, or as a lobbying tool	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Tools for reporting & statistics		
249	4	Collection Manager, Administrator	advertize our collections, and in particular highlight any features of special interest, like presence of type specimens, themed collections, historical collections, named collectors, time series, relevance for specific fields of study	we can develop defendable digitization plans around strategic, fact-based criteria, establish contact with holders of related collections, and attract funding to support our efforts of further digitization, external services and collaboration	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time,taxonomic scope, list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Tools for reporting & statistics		

250	4	Policy maker, Funder	create incentives to publish collection data.	increase the value and use of the collections. Highlight the the relationship between data publishing and data use.	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, possibly list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Tools for reporting & statistics		
251	4	Policy maker, Funder, Collection Manager	publish collection data via data aggregators.	information on biodiversity is expanded	collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, list of taxon / species names, important collectors, thematic interest focus, digitization status	Metadata on collection level	Tools for data analysis	Tools for reporting & statistics		
252	4	Director, Administrator	know what makes our collections unique	can effectively advertize, highlight for use	collection types, with size, locality scope, time, taxonomic scope, important collectors, etc.	Metadata on collection level	Tools for data analysis	Tools for reporting & statistics		
253	4	Policy maker	see how many records exist of a certain group of animals in a particular Geographic area over time not currently vouchered.	can determine the need for Preservation space and document populations over time	collections of species and places	Tools for data discovery	Tools for data visualization	Tools for geo-referencing	Tools for reporting & statistics	Legal and policy framework
254	4	Policy maker	find a list of all e-mail addresses of herbaria in a specific region	general communications can be sent out efficiently	contact data (e-mail addresses) of an herbarium	Metadata on collection level				
255	4	Policy maker	obtain information on the distribution of species under the different nature directives	can assess the conservation status and distribution ranges of protected species	detailed distribution data	Distribution data	Legal and policy framework	Ecological data	Tools for data discovery	
256	4	Policy maker	information on the distribution of species under the nature directives	assess conservation status and distribution range	detailed distribution data	Distribution data	Legal and policy	Ecological data	Tools for data discovery	
257	4	Director	to be able to underline the importance of bio-diversity and of scientific collections in understanding it.	can help policy makers understand the consequences of their policies	details of collections and related research	Legal and policy framework	Tools for reporting & statistics	Tools for data discovery		
258	4	Policy maker, Collection Manager	have a tool to estimate undigitized backlog	hidden collections can be revealed, estimate the remaining effort for digitization	estimating the remaining effort for digitization	Tools for reporting & statistics				
259	4	Policy maker	find information on the contribution of DiSSCo to the international environmental policy agenda	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			
260	4	Policy maker	find and reuse digital specimens from DiSSCo	confirm the presence of a species for legal purposes	fast access to the DiSSCo infrastructure	Legal and policy framework	Tools for reporting & statistics			
261	4	Director, Administrator	document the collections of a cross-disciplinary not-for-profit museum.	my executive team can quickly describe holdings of a particular type/group to an interested trustee/donor on the fly.	geography, classification	Tools for reporting & statistics				

262	4	Association	To gather information to have overall figures representative of partners' state-of-the-art	we can showcase the relevancy of the collections hold to policy makers and attract funds	High-level figures that feature the collections as a whole	Legal and policy framework	Tools for reporting & statistics		
263	4	Conservation Planner	cross-check species identification against reliably identified specimens	create a checklist of species	images, sequence data, georeferences, traits	2D images	Molecular data	Morphological data	Metadata on collection level	Distribution data
264	4	Biomonitoring Planner	cross-check species identification on diatoms against reliably identified specimens	create water quality assesment	indicator values, digitized microscopic collections	2D images	Tools for reporting & statistics		
265	4	Policy maker	know the use of the collections by other domains as a key indicator of its impact	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	Legal and policy framework	Tools for reporting & statistics		
266	4	Director	understand the relationship of our own collection to the collections of other institutions in our country.	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	Legal and policy framework		
267	4	GBIF Node Manager	manage collection metadata in a national scale catalog, while being aware of overlaps and possible conflicts with descriptions held/curated in other places (e.g. IH).	can maintain an overview of our national holdings and can provide relevant and up-to-date information to my constituents, and can spot conflicts with such duplicates to see that they are resolved or highlighted	institutional metadata (name, institution id, physical address, website, contacts); collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, important collectors, thematic interest focus, digitization status; links to related objects (collection metadata entry) in other parts of the catalogue; editing trail (who, when, what action), link assertion (duplicate of, older version of, etc).	Metadata on collection level	Tools for reporting & statistics		
268	4	GBIF Node Manager, Collection Manager	register a core description of collections under my domain.	the existence of the collection is publicly visible next to others, its potential value can be made explicit, and there is motivation for the linking, capture and publication of additional descriptive information (specimen records, checklists, image data, sequences, etc). I can also use this to argue the need for funding of both maintenance and further digitization of the collection itself.	institutional metadata (name, institution id, physical address, website, contacts); collection name, collection id, storage/preservation type, number and kinds of items, geographic scope, time, taxonomic scope, important collectors, thematic interest focus, digitization status.	Metadata on collection level	Tools for reporting & statistics		
269	4	Policy maker	keep track of all specimens collected in National Parks	inventories and collection assesments can be made up	location at which specimens are collected	Advanced search functionality	Distribution data		
270	4	Director	know the extent of the use of the organisation's collections across societal sectors.	can take informed executive decisions related to future investments (both collections and	metrics of the use of collections	Metadata on collection level	Tools for reporting & statistics		

271	4	Researcher, Scientist	easily identify which genetic resources I can directly request from a natural history collection	can fullfill my due diligence obligations under the EU ABS regulation when requesting tissue or DNA-samples from an ex-situ situation inside Europe	not only information on the sample, but also on the status and reference on the access status (without legal doc scan in the www!)	Legal and policy framework	Metadata on collection level	Tools for data discovery		
272	4	Policy maker (law enforcement)	find quality indicators of an herbarium	decision can be made on granting permissions (e.g. CITES permit)	number of type specimens of the collection, total number of specimens, number of researchers attached to the institute, level of digitalisation...	Metadata on collection level	Legal and policy framework			
273	4	Law Enforcement	find back specialists in a specific domain	request identification of unknown specimens (e.g. at customs)	people connected to an herbarium with their respective expertises.	Metadata on collection level	Tools for identification			
274	4	Policy maker, Collection Manager	track changes in collections over time (name changes, merges,)	there is a unique way to reference to collections (e.g. linking specimens to the correct collections after a collection merge or split)	stable collection identifier, collection names	Metadata on collection level	Data integration			
275	4	Policy maker, Collection Manager	know nested relationships between collections	In case of collection merges or splits, it is possible to trace back the origin of a collection	sub-collections	Metadata on collection level				
276	4	Collection Manager, Director, Administrator	know the situation with collection amounts	can plan ahead for future needs for new space and storage	to know existing amounts of collections, and amount of new material coming in. Also, need to know status / condition (wet, dry, ...) of existing material. Also collection health information.	Metadata on collection level	Tools for reporting & statistics			
277	4	Collection Manager	start a digitisation project	can digitize a certain group of my collection; I like to do this internationally because of funding	to know which other institutes hold collections of this group	Metadata on collection level	Tools for data discovery	Data integration		
278	5	Teacher	find out if there is a printable version of a 3D model with a high resolution 3D model	can create a 3D physical model for study in class and can use a hologram projector for inspection of the specimen in class	a reference to the available 3D models	3D images	Tools for data discovery			
279	5	Student	find the referencing information for an image	can cite the source in a written report which includes the image	a reference to the citation information for the image	2D images	Label images	Data integration		
280	5	Student	find a high resolution image	can perform morphological analysis as part of my coursework	a reference to the higher resolution image and the clear procedure to retrieve it	Advanced search functionality	2D images			
281	5	Teacher	to be able to have accurate physical models of animals and plants	can use them in my lessons to illustrate the biology of life-forms	access to taxonomic information, photographs, measurements etc of specimens	Advanced search functionality	2D images	Label images		
282	5	Teacher	show that organisms have restricted geographical distributions	can teach on biodiversity and biogeography	distribution data of organisms	Tools for geo-referencing	Tools for data visualization	Distribution data		
283	5	Student	do autonomous exercises linking multiple types of data	gain a deeper understanding of biodiversity and its evolution; understand the purpose, power and limits of collection data	DNA sequence link, taxon interactions, morphology, location, taxon concept	Molecular data	Morphological data	Distribution data	Ecological data	Data integration

284	5	Teacher	be able to provide accurate scientific information about species and specimens	can educate students/the public about the natural world	easily accessible and up-to-date information about the specimens in museum collections	Advanced search functionality	Tools for data discovery			
285	5	Head, Educational Development	browse all digitized collection data in our institute	could use this for context rich, digital educational activities. In which we not only focus on biological concepts, but also on the cultural historic aspect of our collection and research. This fits well with our goals to enhance science literacy.	field notebooks / diaries, correspondence, (historic) maps drawings, photos, paintings, audio, video	Tools for data discovery	Images related to collections	Data integration		
286	5	Digital Collection Manager, Citizen Scientist	browse all digitized data linked to an expedition	create online and in-house exhibitions focused around expeditions where specimens and cultural objects tell a story of what was collected and what existed along their expedition path, their field notes (from Archives) describe their daily activities in a way which cannot be gleaned through the scientific record of collection and publication and add a humanities angle to the expedition, the literature that arose from specimens found on the expedition can be sourced from the	field notebooks / diaries, correspondence, (Historic) maps,drawings, photos, paintings, audio, video	Tools for data discovery	Images related to collections			
287	5	Student	check my identification of a specimen	can test my bioliteracy and and learn more about a taxon/ecosystem I am studying	image/morphology, links to papers / protologue, collection metadata (habitat, taxon interactions)	Tools for data discovery	2D images	Images related to collections		
288	5	Exhibition Designer	be able to find the location of specimens and information about species.	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				
289	5	Science Communication Officer	know about the research being carried out in natural history collections	can organize a dissemination event	information on research topics and the people behind them	Metadata on collection level				
290	5	Teacher	find out extant relatives of fossil plants in a living collection of a botanic garden	teach palaeontology for my students	list of living plants in collections, links their relative extinct species, digitized specimens of extinct species	Advanced search functionality	Metadata on collection level			
291	5	Teacher	show morphological and genetic diversity within a taxon	can explain basic concepts in evolution, systematics, taxonomy, ecology etc.	overview of comparable data (image, DNA sequence link, location) for multiple specimens from taxonomic group	Molecular data	Morphological data	Distribution data	2D images	
292	5	Teacher	learn about the variety of life and the relationships among species and their environment	teach my students that we are all inter-connected and that Homo sapiens is not the "best" species on Earth	species names, images, distributions, habitats, needs, behaviour, ecological niches, relationships	2D images	Morphological data	Distribution data	Ecological data	

293	5	Curious person	learn about the species that might be in my environment	can improve my bioliteracy	taxonomic fields, common names, geographic coordinates, species characteristics, images	Advanced search functionality					
294	5	Student	be able to identify the species in scientific field trips, without visiting the relative collections of NHMs	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.)	2D images	Morphological data	Distribution data			
295	6	Developer	find a 360 degree view or a 3D model of a specimen	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					
296	6	Publisher	include a digital image from a collection in a scientific paper	can illustrate the publication	access to the digital images and copy rights to use the image	2D images	Tools for downloading data/metadata				
297	6	Digitisation Officer	produce digital specimens from a digitisation line	can upload the images to a customer's CMS	an automated workflow minting PIDs	Tools for uploading					
298	6	Mining Company Official	know where a species occurs based on museum collections and occurences	map the distribution of a wanted mineral through a metallophyte (plant indicating presence of a mineral)	collections digitized, observations, location, date	Advanced search functionality	Distribution data				
299	6	Automatic identification systems developer	Which collections are available to use as a reference (training data set)	can training my algorithms for automatic identification	collections of target species (validated)	Interoperability	Advanced search functionality	Metadata on collection level			
300	6	Collection Manager	look up herbaria that have complementary collections	specimens can be exchanged	collector data, geographical regions of collection items, taxon data	Metadata on collection level	Tools for data discovery	Data integration			
301	6	Collection Manager	look up the correct shipping address of other herbaria	address shipping boxes for loans and other exchanges of specimens	contact metadata of institutions	Metadata on collection level					
302	6	Software developer	develop new usages of the data and ways to add information to the data, through apps or web interactions	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. Addition of geographic locality data as most users hold an handheld GPS device.	detailed collection level information	Metadata on collection level	Tools for data discovery				
303	6	Software developer	train my image recognition software	can use it for taxon recognition	images, digitized specimen data	2D images	Morphological data				
304	6	Systems developer	know information on which collections are available to use as a reference (training data set)	can train my algorithms for automated identification	information on validated collections of target species	Metadata	Tools for identification				

305	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, locality data, physical storage volumes, plus an insight on what is digitally represented and what is not. Institution digitisation priority list.	Metadata on collection level	Tools for data discovery			
306	6	Collection Manager	have a metric on reliability of other herbaria	decide over loan requests	number of type specimens of the collection, total number of specimens, number of researchers attached to the institute, level of digitalisation...	Metadata on collection level	Tools for data discovery			
307	6	Collection Manager	look up individuals with a specific expertise	can request identification or evaluation of a specimen	people attached to instutute and their respective expertise	Metadata on collection level				
308	6	Project Leader	identify participants and resources	the relevant people/institutions are included in the project	possible data elements are: - herbaria that contain type specimens - location of the herbarium - expertise of an herbarium	Advanced search functionality	Metadata on collection level			
309	6	Solution Provider	tap into the vast market of digital storage solutions for digital natural collections	can sell my services and consultancies	predictable numbers on collection type, volume and progress in digitization	Tools for reporting & statistics	Tools for data discovery			
310	6	Solution provider	tap into the vast market of digital storage solutions for digital natural collections	can sell my services and consult	predictable numbers on collection type, volume and progress in digitization	Tools for reporting & statistics	Tools for data discovery			
311	6	Crop Breeder	know where crop wild relatives (phylogenetic related species) are growing	select species that are adapted to certain environmental conditions	to combine phylogenetic information with species occurrence data	Tools for data discovery	Tools for data analysis	Tools for geo-referencing	Distribution data	Ecological data incl. traits
312	6	Editor	use Index Herbariorum codes / unique identifiers of collections	identify the place of deposition of a specimen and specimens can be cited correctly and	unique identifiers of collections IH(-like), specimens connected to a collection	Metadata on collection level	Interoperability	Data integration		
313	6	Solution Provider	build and provide solutions and related services	the keepers and scientists can work better and easier with their collections for less cost	volumes, locations and physical sizes plus an insight on what is digitally represented and what not. Institutions priority as to what needs to be digital first.	Metadata on collection level	Tools for data discovery			
314	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	Tools for data discovery	Tools for data analysis	Tools for downloading data/metadata		
315	7	Saxophonist, Composer, Producer, Educator	map the nucleobase and amino acid sequences that comprise the orchid's DNA codes to form individual movements for each of the five evolutionary families of the genus	compose an étude (Orchidées) on the evolution of the genus Orchidaceae	DNA sequences	Molecular data				
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.

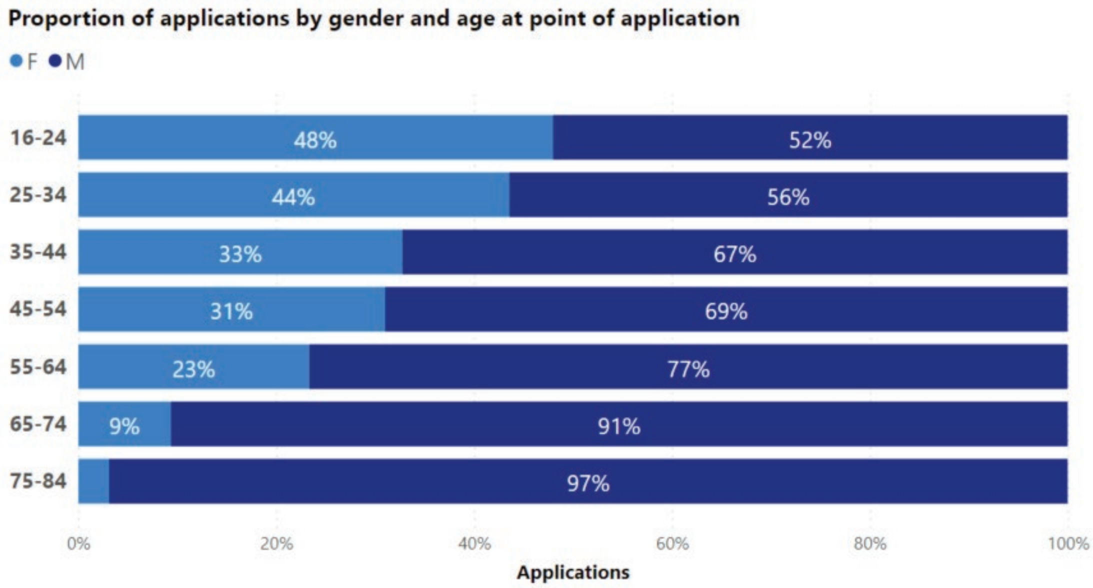

Figure A3.1. Proportion of applications by gender and age at point of application (F - female, M - male).

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.

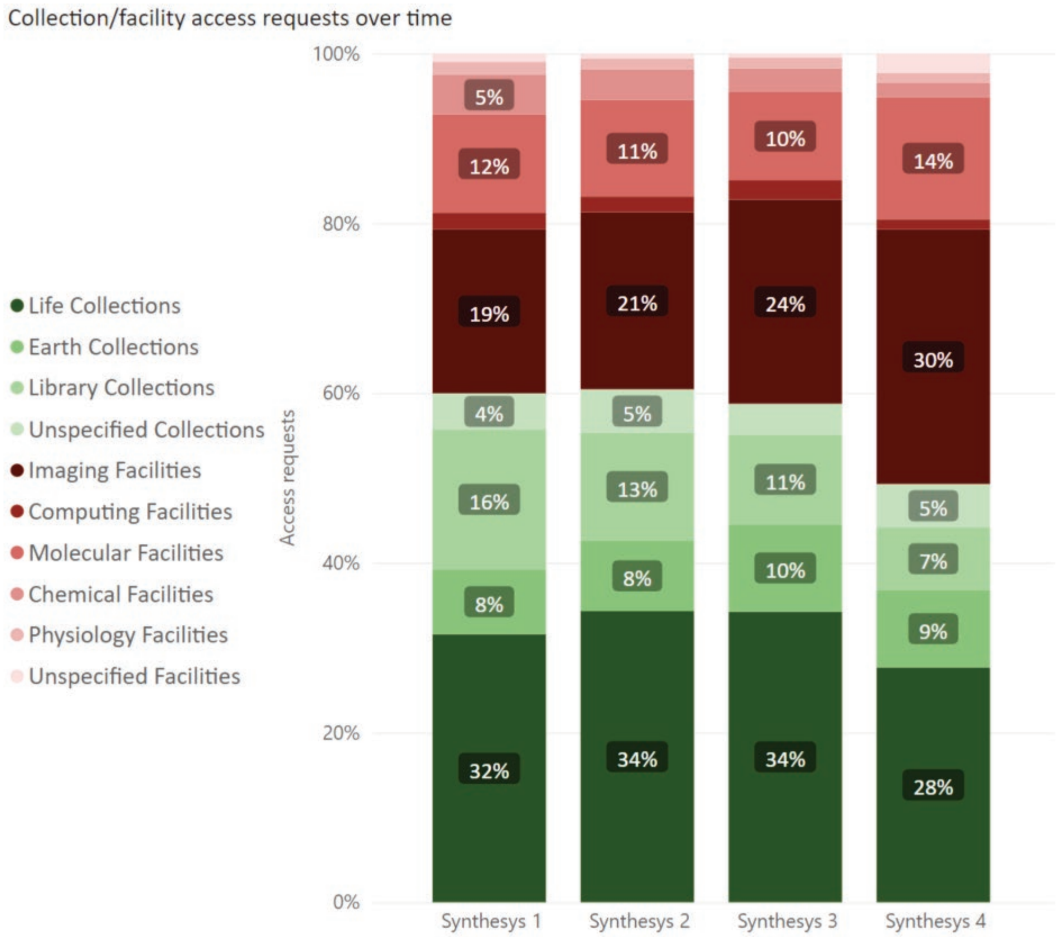

Figure A3.2. Chart showing distribution of request count by category of host facility or collection. Round 1 (n = 8.5k), Round 2 (n = 5k), Round 3 (n = 6.2k), Round 4 (Incomplete: n = 2.8k).

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.

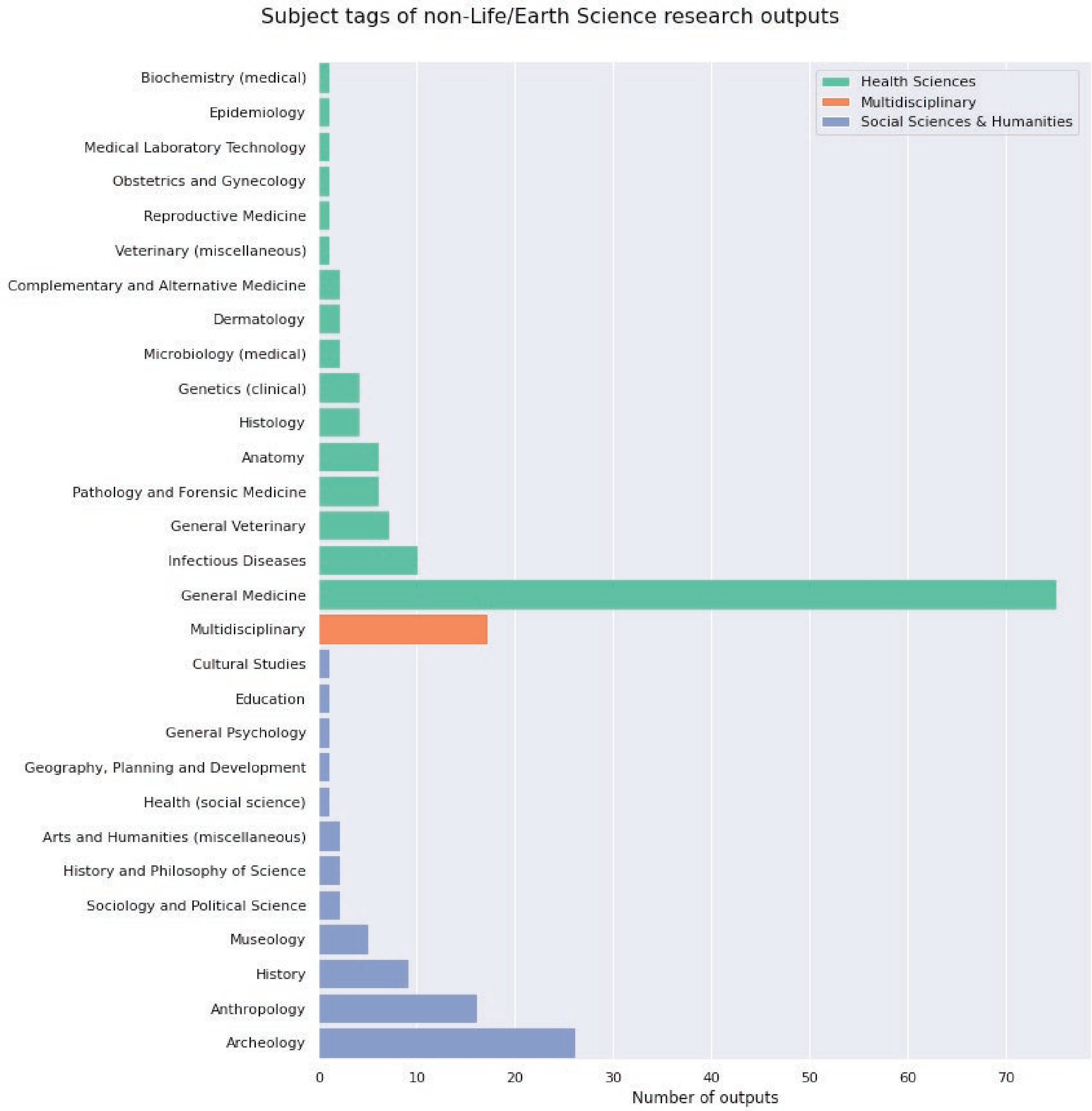

Figure A3.3. Subject tags of non-Life/Earth Sciences research outputs.

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.”