# 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 Milestone report MS 5.1 "Functional technical implementation of DiSSCo Knowledgebase and documentation of most relevant building blocks"


**Author(s)**
Sabine von Mering
Julia Pim Reis
Falko Glöckler
Mathias Dillen
Robert Cubey
Anton Güntsch
Mareike Petersen

**Identifier of the author(s)**
https://orcid.org/0000-0003-2982-7792 (SvM)
https://orcid.org/0000-0003-4383-0357 (JPR)
https://orcid.org/0000-0002-7127-2738 (FG)
https://orcid.org/0000-0002-3973-1252 (MD)
https://orcid.org/0000-0001-7902-3843 (RC)
https://orcid.org/0000-0002-4325-4030 (AG)
https://orcid.org/0000-0001-8666-1931 (MP)

**Affiliation**
Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science

**Contributors**
Wouter Addink https://orcid.org/0000-0002-3090-1761
Alex Hardisty https://orcid.org/0000-0002-0767-4310
David Fichtmüller https://orcid.org/0000-0002-0829-5849
Kessy Abarenkov https://orcid.org/0000-0001-5526-4845
Matt Woodburn https://orcid.org/0000-0001-6496-1423
Elspeth Haston https://orcid.org/0000-0001-9144-2848



**Publisher**
DiSSCo Prepare

**Identifier of the publisher**

**Resource ID**
https://doi.org/10.34960/skvy-bq35

**Publication year**
2021

**Related identifiers**

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

**Creation date**
15/07/2021

**Version**
1

**Citation**
von Mering S., Pim Reis J., Glöckler F., Dillen M., Cubey R., Güntsch A. & Petersen M. (2021): DiSSCo Prepare Milestone report MS 5.1 "Functional technical implementation of DiSSCo Knowledgebase and

documentation of most relevant building blocks". https://doi.org/10.34960/skvy-bq35

**Abstract**
The DiSSCo Prepare Milestone report MS5.1 “Functional technical implementation of DiSSCo Knowledgebase and documentation of most relevant building blocks” describes the approach taken in developing the Knowledgebase (KB) as a central hub for research outputs and technical documentation related to DiSSCo. Information types to be covered in the KB and potential software components are described. The DSpace system was chosen as a central document repository and is available in a beta version at https://know.dissco.eu/. Feedback from project partners will be prioritized to decide on the next steps in further development of a knowledge hub for information on DiSSCo-related topics.

**Content keywords**
scientific

**Project reference**
DiSSCo Prepare (GA-871043)

**WP number**
WP5

**Project output**
Milestone report

**Deliverable/milestone number**
MS5.1

**Dissemination level**
Public

**Rights**

**License**
Attribution 4.0 International (CC BY 4.0)

**Resource type**
Text

**Format**
PDF

**Funding Programme**
H2020-INFRADEV-2019-2

**Contact email**
sabine.vonmering@mfn.berlin

DiSSCo PREPARE

H2020-INFRADEV-2019-2
Grant Agreement No 871043

# DiSSCo Prepare WP5 – Milestone report

# MS5.1 Functional technical implementation of DiSSCo Knowledgebase and documentation of most relevant building blocks

Work package lead: Mareike Petersen (MfN)

Authors: Sabine von Mering (MfN), Julia Pim Reis (MfN), Falko Glöckler (MfN), Mathias Dillen (MeiseBG), Robert Cubey (RBGE), Anton Güntsch (BGBM), Mareike Petersen (MfN)

Contributors: Wouter Addink (Naturalis), Alex Hardisty (U Cardiff), David Fichtmüller (BGBM), Kessy Abarenkov (U Tartu), Matt Woodburn (NHM), Elspeth Haston (RBGE)

MUSEUM FÜR NATURKUNDE BERLIN

# Abstract

The DiSSCo Prepare Milestone report MS5.1 “Functional technical implementation of DiSSCo Knowledgebase and documentation of most relevant building blocks” describes the approach taken in developing the Knowledgebase (KB) as a central hub for research outputs and technical documentation related to DiSSCo. Information types to be covered in the KB and potential software components are described. The DSpace system was chosen as a central document repository and is available in a beta version at https://know.dissco.eu/. Feedback from project partners will be prioritized to decide on the next steps in further development of a knowledge hub for information on DiSSCo-related topics.



# Keywords

DiSSCo, DSpace, FAIR data, information transfer, knowledge base, knowledge hub, repository, system architecture

# Index

# 01 INTRODUCTION

As an initiative formed by public research institutions, the Distributed System of Scientific Collections (DiSSCo) is committed to Open Science. Open Science not only makes scientific work more transparent and accessible, but also enables a whole new set of collaborative and IT-based scientific methods. Therefore, the outputs of our common research projects should be openly available as much as possible and research data easily Findable, Accessible, Interoperable and Reusable (FAIR principles).

DiSSCo Prepare (DPP), the preparatory phase of DiSSCo, is building on profound technical knowledge from various sources and initiatives. Efficient knowledge and technology transfer for partners building the DiSSCo technical backbone will be facilitated by a central and freely accessible DiSSCo Knowledgebase, designed and implemented within the Work Package "Common Resources and Standards" and specifically task 5.1 "DiSSCo Knowledgebase for technical development". As a hub for knowledge management relevant within the DiSSCo context, the DiSSCo Knowledgebase (KB) will not only store all research outputs from DiSSCo-linked projects and other resources in one place but also further building blocks relevant for users of the DiSSCo Research Infrastructure (RI). Such building blocks include web services, PID (persistent identifier) systems, controlled vocabularies, ontologies and data standards for bio-and geo-collections objects, collection descriptions, digital assets standards as well as domain-specific software for quality assurance and monitoring.

## Approach

### Information types to be covered in the Knowledgebase

In close collaboration and exchange with other DPP project partners, the task group collected the extent of information types expected to be stored in the knowledgebase. To get a more complete picture, this was also discussed together with project overarching bodies such as the DiSSCo Technical Team. As a last preparatory step, a survey was sent to all task and work package leads of DPP to evaluate which information types partners are planning to make available via the knowledgebase. The feedback was included in the discussions and next planning steps. The latest overview of desired information types is given in Figure 1.


**Fig. 1:** Information Types in the DiSSCo Knowledgebase. This expected cluster of information categories (blue dots) was based on DPP project outcomes and relevant external resources.

As the term Knowledgebase traditionally was used in a context of providing machines with a database of facts for reasoning processes, the partners agreed that we would use the term with a main focus on human readability in the DiSSCo Knowledgebase in the first place. The importance of machine readability varies amongst different information types. However, the metadata will be machine-readable in a consistent manner.

**Landscape analysis to select a system as document repository**

A comprehensive landscape analysis with short presentations of each system took place during two task group meetings. The following candidate systems were introduced by different task members: DSpace + Vivo, Alfresco, Fedora, OSF, Liferay, Invenio, Dataverse, and E-Prints. For the decision process, requirements of the knowledgebase were collected and prioritised.

Criteria of top priority for the decision of an appropriate component for the knowledgebase to serve the information type “Public Documents and External Resources” were:

- Capability of storing documents and free text for referencing deliverables, publications and Questions and Answers / FAQs
- Extensibility & customization (plugins or extensions)
- Comprehensive public technical documentation and user documentation
- Comprehensive REST API
- Mechanisms for stable versioning of content
- Search index (including the capability of indexing of customizable metadata)
- Hierarchical structuring of pages and other entities
- Capability of structuring the content by categories, tags or labels
- File upload, storage and download
- User-friendly search functionality
- Regular security updates

- View and download functionality for common document and image file formats
- Option to run an instance in a cloud environment (rather than a Software as a Service approach)
- Sustainability of the software product (e.g. organisation in place to support and maintain)

Based on the requirements, the most promising systems were DSpace, CKAN, and Alfresco. All three products meet the requirements for the respective information type "Public documents and external resources" in the knowledgebase according to the prioritized criteria. So, the following additional aspects with respect to the implementation and maintenance have been included in the decision process: latest releases, size of user community, regular support and good software maintenance allowing the correction of possible bugs, and regular security updates. Thus, the team chose DSpace, an open source repository software package of rich and powerful features that focus on long-term storage, access and preservation of digital content. It is available as free software under an open-source license in a public GitHub repository and has a huge user community and a very active group of developers. It offers customizable interfaces, a full-text-search where the provided metadata for content is indexed to be searchable and accessible with the use of a REST API enabling the data to be FAIR. A reliable search functionality allows the end-users to find the content without delay even for huge amounts of data which is essential regarding scalability with an increasing amount of linked information. A list of more convincing key features of DSpace can be accessed at the official website.

The approach and decision process was also presented to a wider audience in a blog post published in December 2020 in the DiSSCo Tech blog.

**Working session at the DiSSCo Prepare All Hands Meeting**

A first beta version (see Fig. 2) was made available and introduced during the First virtual All Hands Meeting of DiSSCo Prepare (January 18-22, 2021). This event brought together project partners, with the objective to present and discuss what will become Europe's leading natural science collections Research Infrastructure, the DiSSCo RI. In a dedicated section, the DiSSCo Knowledgebase and its functionalities were presented to the audience in order to discuss possibilities to structure the content, and to present a number of features such as the full text search functionality, possibility of different types of file upload, metadata customization, etc.
The participants could test the first version by browsing the software and testing features and tools, allowing feedback and requirements from DiSSCo partners.

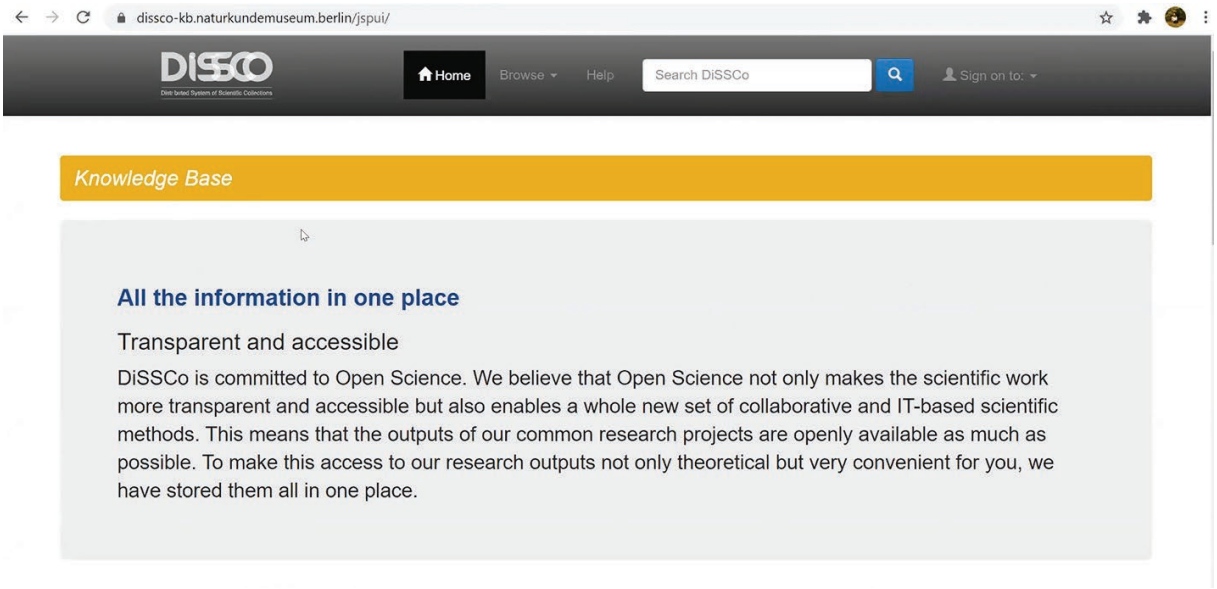



**Fig. 2:** Screenshot of the first beta version of DiSSCo Knowledgebase homepage.

Some feedback on the first beta version was collected directly during the meeting. A dedicated GitHub repository (see Fig. 3) was set up to collect all issues and feedback related to the Knowledgebase.

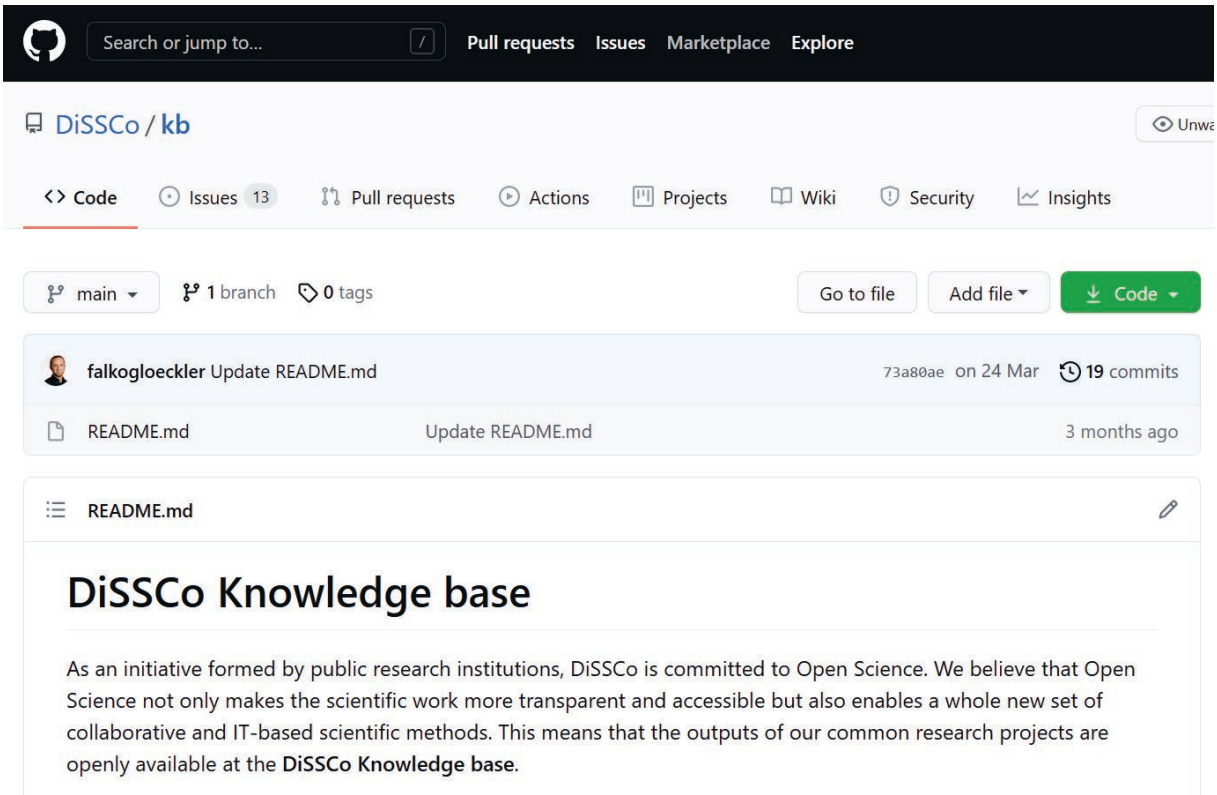



**Fig. 3:** Screenshot of GitHub repository for the DiSSCo KB (https://github.com/DiSSCo/kb).

**DiSSCo Prepare Round table on "Organisation of knowledge and documentation for stakeholders"**

The virtual DiSSCo Prepare Round table "Organisation of knowledge and documentation for stakeholders" took place on July 6th, 2021 with more than 60 participants from the DiSSCo Prepare project and other interested parties. The goal of the meeting was to identify the stakeholder groups that need access to DiSSCo knowledge and documentation, to discuss how to organise the knowledge in such a way that optimal access is provided for these stakeholders, and to identify resources (existing elsewhere or missing) that should be provided.

The round table started with an introduction to the DiSSCo Knowledgebase (see presentation), including expected information types and potential software components. Mentimeter was used to collect information on the participants and to get feedback on questions such as *Where are you from? How are you involved [in the project]? Why are you attending today?* The answers showed that most attendees were from Europe and team members of DPP. People were mostly attending to find out more about the KB and to contribute resources. Additionally, a Mentimeter survey was used to collect feedback on existing sources and added value of the DiSSCo KB (see Appendix 1).

After the surveys and a general discussion, five stakeholder groups were proposed as the topic of the break out groups:

1. Collection staff (curators, collections managers & data managers, etc.)
2. Researchers
3. Developers
4. Policy makers & funders
5. Citizens, citizen scientists & wider public

While initially the focus was on internal stakeholders, the Knowledgebase can be extended for external users as well and should reach out to their needs. A detailed account of the DiSSCo Prepare Round table including the slides of the presentations, the results of the Mentimeter survey and the most important outcomes of the breakout groups is provided in Appendix 1.

## Content of the DiSSCo Knowledgebase

According to our findings, the different information types vary in formats and system requirements and cannot be stored in one single system. Whereas for some information types the target system is more or less set (e.g. GitHub for software code), for others a well considered decision is necessary. Task partners focused on a decision about a software system for the most common and required information type "Public Documents and External Resources" in order to aggregate references to distributed documents and sources in a single point of entry.

The Knowledgebase will provide:

(1) structured technical documentation of identified DiSSCo technical building blocks, such as web services, PID systems, controlled vocabularies, ontologies and data standards for bio- and geo-collection objects, collection descriptions, digital assets standards as well as domain-specific software products for quality assurance and monitoring,

(2) an assessment of their technical readiness for DiSSCo, and

(3) specifications on their relevance for the overall DiSSCo technical infrastructure and the DiSSCo data model (compare to MS5.6 "A functional prototype of DiSSCo Modelling Framework" of Task 5.2).

**DSpace as a document repository**

As a central requirement of the Knowledgebase, the DSpace system will function as a document repository storing all kinds of different content types. “Public documents & external resources” were identified as the most important information type.

Below a welcome text block on the KB homepage, buttons that deep link to knowledge sections allow easy access to different highlighted resources. The KB homepage will function as a single entry point to different DiSSCo components (see Fig. 4).

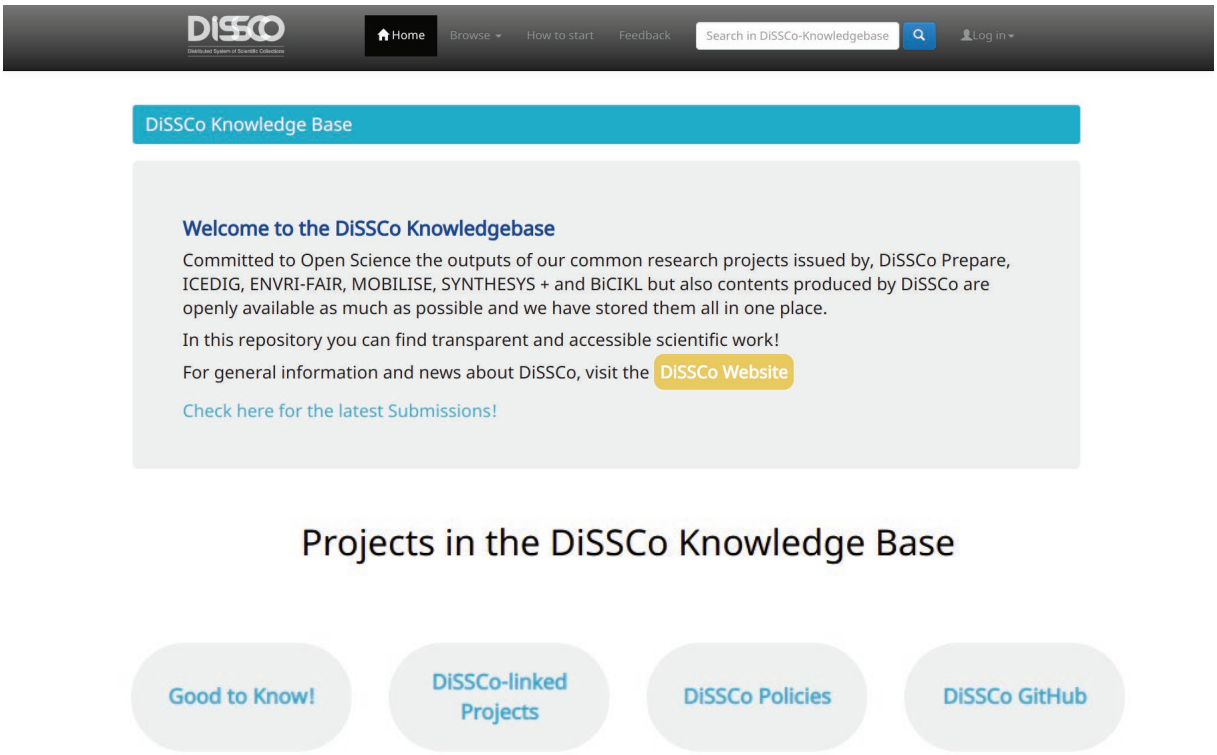



**Fig. 4:** Screenshot of the highlighted projects in the homepage, by clicking on them the user will get an overview of the different information types, either in the DSpace instance but also in external sources (e.g. Github).

Collections of resources in the DSpace instance are organised in different levels, named ‘projects’ and ‘clusters’. The most important ones are listed below:

1) **Good to Know!** - a compilation of material providing more general information on DiSSCo and related topics. Currently, it comprises the following thematic clusters:
    - **FAQs:** Frequently asked questions & their answers,
    - **Glossary:** a list of terms and short descriptions and links related to and relevant for DiSSCo.

2) **DiSSCo-linked projects:** collection of project outcomes such as Milestone reports and Deliverables, i.e. output of DiSSCo-linked projects such as BiCIKL, DiSSCo Prepare, ENVRI-FAIR, ICEDIG (see Fig. 5), MOBILISE, and SYNTHESYS+. Within the project clusters, documents are sorted by the projects' work packages.

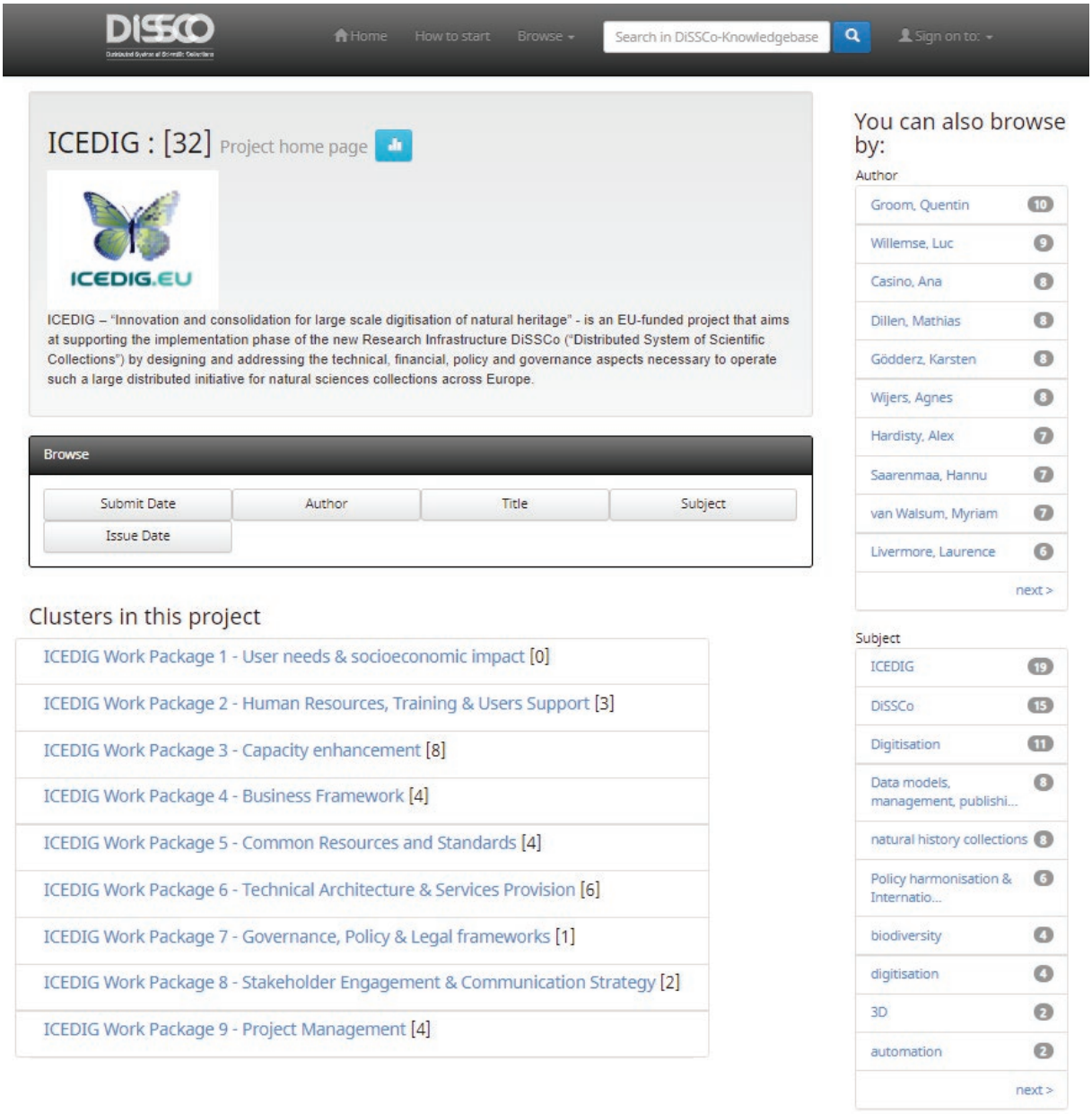



**Fig. 5:** Screenshot of the DSpace 'cluster' compiling output resources from all work packages of the DiSSCo-linked project ICEDIG (https://know.dissco.eu/handle/123456789/4). Filtering options are shown on the right side.

3) **DiSSCo Tech:** information and technical documentation related to the DiSSCo technical architecture and all technical developments:
    - technical documentation (e.g., system documentation, guidelines for both internal and external developers),

- **DiSSCo Tech Blog** posts: a collection PDF representations of the blog posts originally published at DiSSCo Tech,
- information on **data standards, ontologies**, etc. (the information compiled in Milestone report MS5.4 “Data standards relevant for DiSSCo” will be made available in the KB),
- information on **PID systems** (the information from Milestone report MS5.3 “Documentation of PIDs relevant for DiSSCo technical infrastructure” due in October 2021 will be used),
- **controlled vocabularies** (e.g. provided in Wikibase) could be linked.

4) **Other resources:** additional DSpace ‘projects’ could provide material and documents around topics such as:
    - **DiSSCo Policy:** collection of relevant policies, policy framework, etc.,
    - **Training materials** relevant for the DiSSCo community,
    - etc.

This initial structure can be easily adapted according to the wishes and requirements of the DiSSCo project partners and the wider community. Once the integration with the other components is implemented, these will be prominently featured on the KB homepage.

## Other information types to be integrated

Eventually, the DiSSCo Knowledgebase will become a single point of entry for many different information types (see Fig. 1). For several such information types, components are already available and need to be integrated (see Fig. 6). An important step is to ensure interoperability between the different components. A connection between the components could be realized by simply linking to other systems via the Knowledgebase homepage (https://know.dissco.eu/) or by a closer integration of used systems or platforms wherever useful and technically feasible.

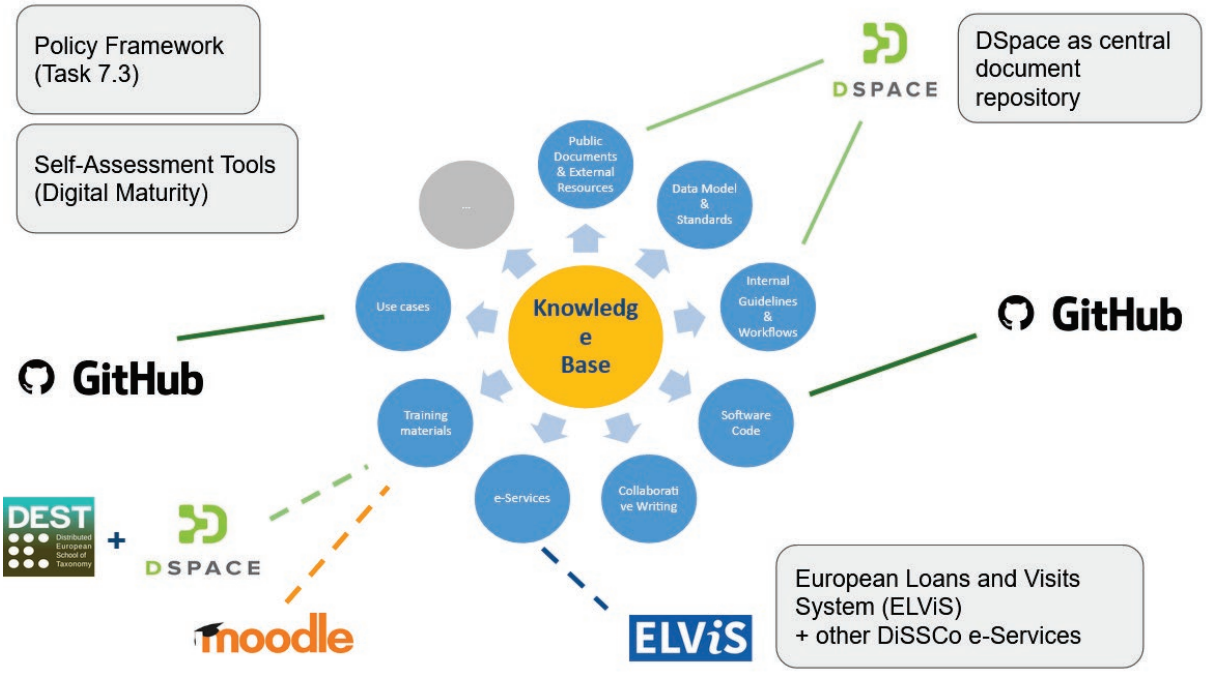



**Fig. 6:** Information types and (potential) software components that comprise the Knowledgebase. Dashed lines indicate the potential tools not implemented yet.

**Software code**

For software code and all related issues, the natural decision is to use GitHub, a commonly used code repository for software development. GitHub offers the features such as version control and source code management (SCM) functionality of Git, access control, and collaboration features (bug tracking, feature requests, task management, continuous integration and wikis for every project). The system is also used by other DiSSCo linked projects, EU projects and other relevant open source systems. DiSSCo's GitHub project is available at https://github.com/DiSSCo and currently comprises 20 repositories focussing on topics such as the European Loans and Visit System (ELViS), the open Digital Specimen (openDS) specification, user stories, etc. (see Fig. 7).

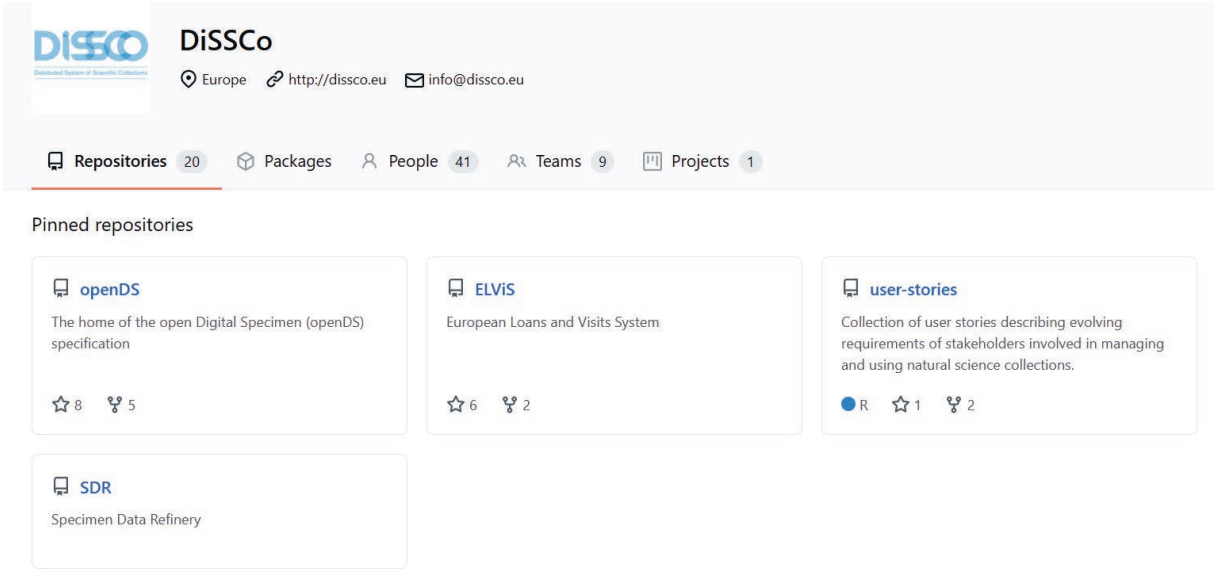



**Fig. 7:** Screenshot of DiSSCo GitHub account with currently 20 different repositories.

**User stories & use cases**

User stories and use cases have been collected by the DiSSCo Coordination team to better understand the needs and requirements of the community and to prepare the development roadmap. Several DiSSCo linked projects have contributed user stories to the collection (e.g. ICEDIG and DiSSCo Prepare) which can be found in the GitHub repository at: https://github.com/DiSSCo/user-stories.

**Training materials**

DiSSCo partners agree that dedicated training is crucial for capacity building in our community in order to achieve the desired readiness levels. It is important that institutions continue to open up and share their expertise and learn from each other. The DiSSCo Knowledgebase will therefore compile and provide access to training material of all kinds. In addition, learning platforms such as Moodle can be integrated. A Moodle plug-in for DSpace is available and might prove very helpful in this respect. Moodle is a free, open-source learning management system widely used for different forms of e-learning.

The Distributed European School of Taxonomy (DEST), established during the CETAF initiative EDIT, offers education and training opportunities in the field of taxonomy, biodiversity, geodiversity, and conservation. DEST activities help to transfer knowledge between current and future generations of taxonomists by providing high quality education. The DEST website provides access to information on training programmes and opportunities in different areas related to e.g. taxonomic research and natural science collections. Possibilities on how to connect this website with the DiSSCo KB will be discussed with project partners in WP2.

**DiSSCo e-Services**

The DiSSCo e-Services are web services developed for or related to DiSSCo. Such e-Services include: the European Loans and Visits System (ELViS), that is already in use within the project SYNTHESYS+, the Digital Specimen Repository (nsidr.org), the Collection Descriptions Dashboard (CDD), the Specimen Data Refinery (SDR), the Helpdesk (initially for ELViS), the Authorisation and Authentication Infrastructure (AAI), the Unified Curation and Annotation Service (UCAS) and the Policy Self-assessment tool. As these services are populated with individual documentations, user guidelines, API specifications, etc. We summarize them as a separate information type in the context of the DiSSCo Knowledgebase.

DiSSCo Labs provides a preview of experimental e-services and demonstrators created by the DiSSCo community.

## Pilot implementation of DiSSCo Knowledgebase using DSpace as document repository

The DSpace installation provides the central entry point of the DiSSCo Knowledgebase and functions as content management system for the information type of regular documents, metadata, frequently asked questions, links to external resources and to other components. At the date of this milestone report, a *beta* version of the DiSSCo Knowledgebase is already available at https://know.dissco.eu/ to provide a pre-release of the desired central and freely accessible hub for knowledge management that will provide unified access to research outputs, technical documentation. This will allow efficient knowledge and technology transfer relevant within the DiSSCo context.

**DSpace Durable Digital Depository**

DSpace offers customizable interfaces, a full-text-search where the provided metadata for content is indexed to be searchable and accessible with the use of a REST API enabling the data to be FAIR. With the reliable search functionality the end-users can find and browse the desired content. DSpace allows storing different versions of documents, adding metadata and free text, hierarchical structuring of pages and stable resource URLs. This enables the DiSSCo partners to reference their content like deliverables, publications and Questions and Answers (i.e. FAQs). Furthermore, an editorial workflow modelled and role based access management in the system allow the project coordinators and administrators to prepare content privately and to review the content which helps to conduct a profound quality assurance prior to publication.

DSpace was developed to be open source, and in such a way that institutions and organizations with minimal resources could run it. The system is designed to run on the UNIX platform, and comprises other open source middleware and tools, and programs written by the DSpace team. All original code is in the Java programming language and requires an open **Java development kit (OpenJDK)**. Other pieces of the technology stack include a relational **database management system (PostgreSQL)** to store and provide access to data points that are related to one another. More information on how to install DSpace is provided at: https://wiki.lyrasis.org/display/DSDOC6x/Installing+DSpace.
In order to enable developers to easily pack, ship, and run DSpace as a lightweight, portable, self-sufficient container, which can run virtually anywhere, the open OS-level virtualization software Docker was used.

**DSpace System Architecture**

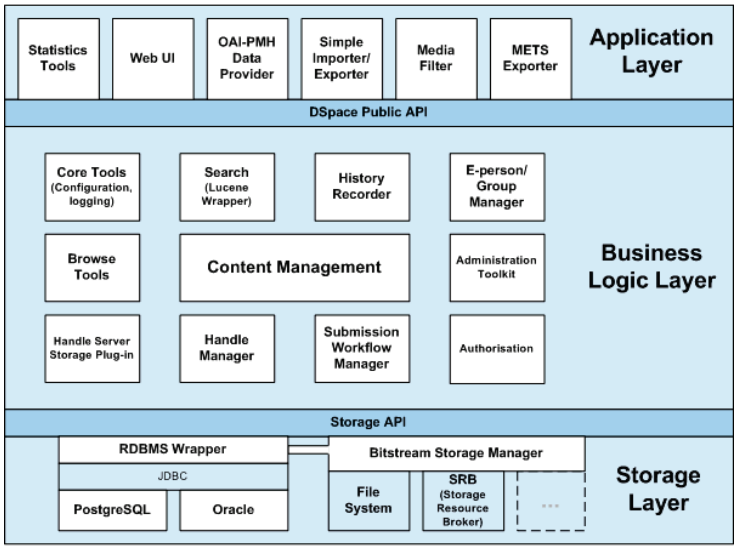



**Fig. 8:** DSpace technical architecture (https://wiki.lyrasis.org/display/DSDOC6x/Architecture)

The DSpace technical architecture is designed to have a three-layer architecture (see Fig. 8), defined by: storage, business, and application layers. A documented API allows for future customization and enhancement. The storage layer is responsible for physical storage of metadata and content,

implemented using the file system, as managed by PostgreSQL database tables. The business layer is where the management of the content, users (e-person), authorization and workflow resides. The application layer contains components for cross-reference that allows DSpace to communicate with external resources, such as Web user interface and the Open Archives Initiative (https://guidelines.openaire.eu/en/latest/) protocol for metadata harvesting service. Each module has an API to allow DSpace adopters to replace or enhance that function as desired (Naik & Naik 2019[1]).

## First Beta version of the DiSSCo Knowledgebase

The online presentation of the content is organized in a flat hierarchy of two levels that are called 'Projects' and 'Clusters', of which the latter represent the lowest aggregation level of content items. Users can access landing pages for individual content items using the full-text search, the faceted browsing, and through external reference such as a persistent identifier (e.g. DOI) in order to find the content even without the hierarchical structure. Thus, multiple approaches to exploring the content are covered.

DSpace can accommodate any type of files uploaded to the system. Within the context of the DiSSCo Knowledgebase the only most common document types will be relevant (e.g., XML, XSD, PDF, XLS, PPT, JPEG), but there is no actual limitation. A lot of content is expected to be submitted in free text format (for FAQs and Glossary) instead of using the file upload. This is beneficial for the full text search, which is limited in the case of certain file types.

The first beta version of the Knowledgebase uses the Metadata Schema (Dublin Core) with a default submission form, and metadata display. The items are displayed in a simple and full metadata item record. However, multiple metadata schemas can be configured and required metadata fields selected from a mix of configured schemas to describe your items.

### Configurable workflow and content curation when uploading an item

Workflows allow submissions to be checked before entering the items into the repository (Fig. 9). This can be achieved by responsible e-persons. An e-person is a user who has permission to edit and administer the cluster). If required, e-persons or a group of e-persons can be assigned to the clusters to check for accuracy, in order to improve the metadata, or simply to decide if the content is suitable to be archived.

[1] Naik P. G. & Naik G. R. 2019. Creating and Managing Institutional Repository Using DSpace - A Case Study Approach. Educreation Publishing, Kolhapur.

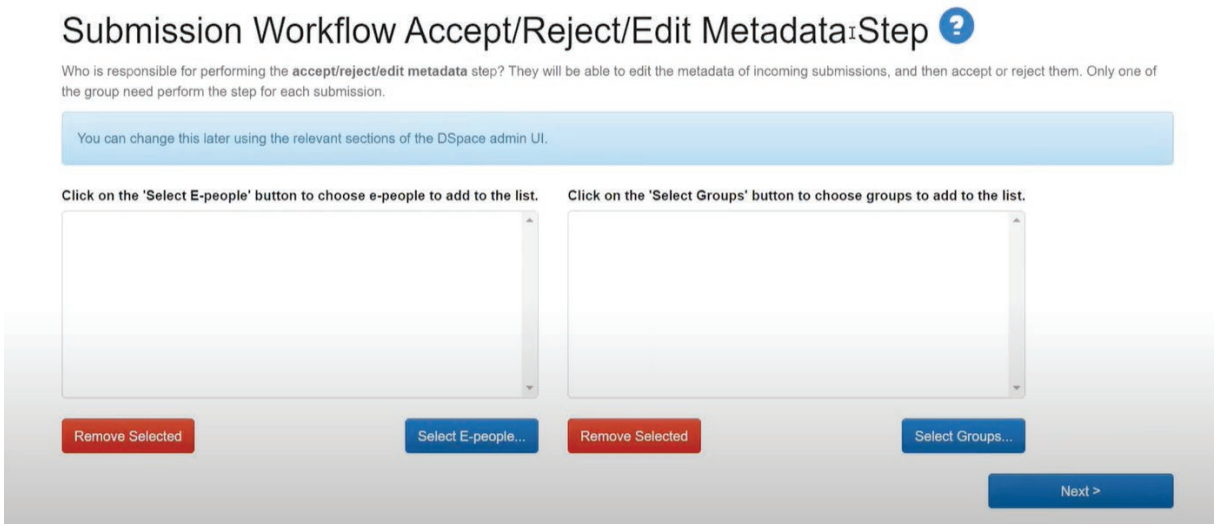



**Fig. 9:** Screenshot of DiSSCo Knowledgebase submission workflow for a specific cluster.

With the feedback collected from the Github repository and the notes from the All Hands Meeting, a number of improvements and adjustments were implemented into the Knowledgebase. Besides layout changes (colours replaced to be similar to the DiSSCo Prepare branding), more general information about DiSSCo and a section dedicated to redirect the user to the DiSSCo website were included in the homepage.

The mandatory metadata fields required by the DiSSCo project output form used to generate the report cover pages (currently at https://www.cognitoforms.com/DiSSCo1/DiSSCoRelatedOutput) were added, including a metadata field for the item's DOI (accessible by full item record view), the publisher, citation and also the abstract to increase the searchability / usability (see Fig. 11).

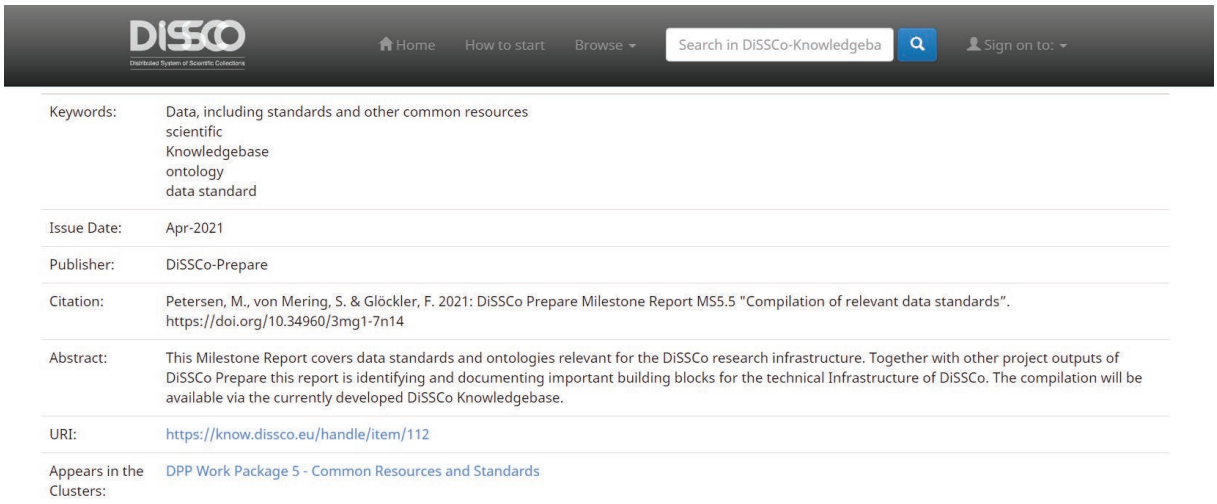



**Fig. 10:** Screenshot of the item's display and its related metadata with the additional fields: Keywords, Publisher, Abstract, Citation (https://know.dissco.eu/handle/item/112).

After applying the suggested enhancements collected by project partners to the Knowledgebase, it is necessary to make a deployment in a production server, this step allows making the software product available to the end-user. To ensure that the product is well tested prior to its release, the following workflow is applied: firstly, the DiSSCo Knowledgebase is tested in the development test environment and re-tested at the pre-production server. At this stage, issues are collected from the partners within the GitHub repository (see Fig. 3) dedicated to the Knowledgebase. The last step is to make the product available on the production server, where it is being hosted and accessed by the end-users.

## Outlook & next steps

The DiSSCo Knowledgebase, even though still in beta version and under active development, has already been useful in providing access to research outputs from various DiSSCo-linked projects. Discussions during the DPP Round Table showed that there is consensus among project partners that linking different resources and components should have priority over developing new functionalities within one system.

Next steps will include further exchange and collaboration with several DPP work packages to discuss and specify further integration of tools and e-services. This will be especially important for the following topics (and the WPs and task groups working on them):

- training materials, integration with learning platform and DEST portal offering training courses and workshops (WP 2)
- the Digital maturity self-assessment tool (WP3)
- Policy tool (WP 7, Task 7.3)

The extensive feedback and ideas contributed during the DiSSCo Prepare Round table meeting will be sorted, structured and prioritized. A single point of entry will be created where project partners (and potentially also external users) could suggest resources for inclusion in the KB. This should be done in a structured way (e.g. in GitHub), by entering the title, the URL and maybe some additional information on the resource.

In the remaining months before the due date of the deliverable, the work on the DiSSCo Knowledgebase will focus on both, essential aspects towards a robust operation of the system, and additional features that have been identified as high priority. For the robust operation the maintenance and update workflows will be revisited and consolidated. Which includes efficient backup/restore procedures, features for web analytics and technical user support according to basic service levels. The basic support would be available at least until the end of the project. The top priority tasks that have been requested by the community comprise the improvement of the user experience and web design (according to the DiSSCo/ELViS style guide), facilitation of importing existing knowledge, including deep- and cross-linking of content (e.g. via ORCIDs), well-established workflows for content managers and contributing users (e.g. automated DOI assignment). Some high (but not top) priority features depend on the progress of other tasks or even external aspects. One example would be the progress on a DiSSCo wide authentication service the Knowledgebase would connect to for Single-Sign-On authentication of users. Another example would concern feature requests regarding DSpace's API layer which is limited in DSpace 6, but full-fledged in DSpace 7. However, DSpace 7 is currently in a Beta version, with active development and testing happening within the community, so that the first release of the DiSSCo Knowledgebase will still be based on the stable DSpace version 6.

The participants of the Round Table expressed the demand to include only well curated content in the KB. To ensure high data quality and consistency, an editorial board for the DiSSCo KB could be established. This needs to be discussed and evaluated among project partners but might help to develop the KB into a trusted resource for the DiSSCo community but also for (potential) external users.

## Appendix 1:

Minutes of the DiSSCo Prepare Round table on “Organisation of knowledge and documentation for stakeholders” on July 6th, 2021.

## Appendix 1: Minutes of the DiSSCo Prepare Round table on "Organisation of knowledge and documentation for stakeholders"

DiSSCo Prepare Round table

# Organisation of knowledge and documentation for stakeholders

**When:** July 6, 2021 – Thursday, 9:00-12:00 CEST (virtual meeting)
**Chairs:** Mareike Petersen, MfN & Wouter Addink, Naturalis

## Introduction

DiSSCo aims to provide open access to the knowledge and documentation that is being produced e.g. in the DiSSCo-linked projects. Through DiSSCo Prepare (DPP) a Knowledge Base is being created: https://know.dissco.eu, a repository that will offer a central place to store all DiSSCo knowledge to make this publicly accessible for DiSSCo stakeholders and anybody interested in this information. The Knowledgebase content is initially organized around Projects which are represented by EC projects in which DiSSCo outputs are being created (DiSSCo Prepare, ICEDIG, ENVRI-FAIR, MOBILISE, SYNTHESYS+, BiCIKL). DPP partners have indicated through a survey which information types they may want to provide through the knowledgebase, these were clustered in groups (see Fig. 1).


**Fig. 1:** Possible information types in the DiSSCo Knowledgebase.

## Round table topics

1. To improve this organisation of knowledge further, a discussion is needed around the potential stakeholders that need access to the knowledge and documentation, to discuss how the information should be organised for each of the stakeholder groups to make it as accessible as possible.
2. Another thing that needs to be discussed is where to make the documentation available/ provide links. Should access be provided through the Biodiversity Knowledge Centre, through GBIF, through some of the other DiSSCo eServices like the Helpdesk, ELViS, or the Policy Framework and Self Assessment tools?
3. The selected solution (DSpace) for the knowledgebase is not suitable for all information types collected through the survey but focuses on the most common information type "Public Documents and External Resources" in order to aggregate references to distributed documents and sources in a single point of entry. For Software code, user stories and use cases DiSSCo is using GitHub. Further discussion is needed to decide on systems for other information types.
4. Know.dissco.eu is not a CoreTrustSeal certified data repository for long term preservation of outputs and it would be hard to get such certification. So for this purpose the outputs may also need to be stored elsewhere, like in Zenodo. What outputs need to be preserved and where? Should we make use of the new Open Research Europe platform for Horizon2020 outputs?

## Goal

The goal of this round table is to identify the stakeholder groups that need access to DiSSCo knowledge and documentation, to discuss how to organise the knowledge in such a way that optimum access is provided for these stakeholders, and to identify resources (existing elsewhere or missing) that should be provided.

## Expected outcomes

- A list of identified stakeholders
- Directions for organising the knowledge
- An initial list of resources to include per stakeholder
- Optional: identification of groups of volunteers who will collect and add identified resources that already exist to the knowledge base

## Agenda

- Round of introductions - 15 min
- General presentation - setting the scene
    - Introduction - 5 min
    - DiSSCo Knowledge Base - 10 min
- Identification of stakeholder groups - 10 min
- Discussion on optimal knowledge organisation for each stakeholder group - 20 min

- Discussion on places where to make knowledge available beyond the knowledge base - 20 min
- Break - 15 min

- Introduction to the breakout session - 5 min
    - Breakout working session in 5 subgroups to identify documentation and outputs for identified stakeholders that need to go into the knowledge base (both existing and not yet existing documentation). 50 min
    - Presentation of subgroup results. 20 min (5x5)
- Identifying next steps and closing - 10 min

## Background reading material:

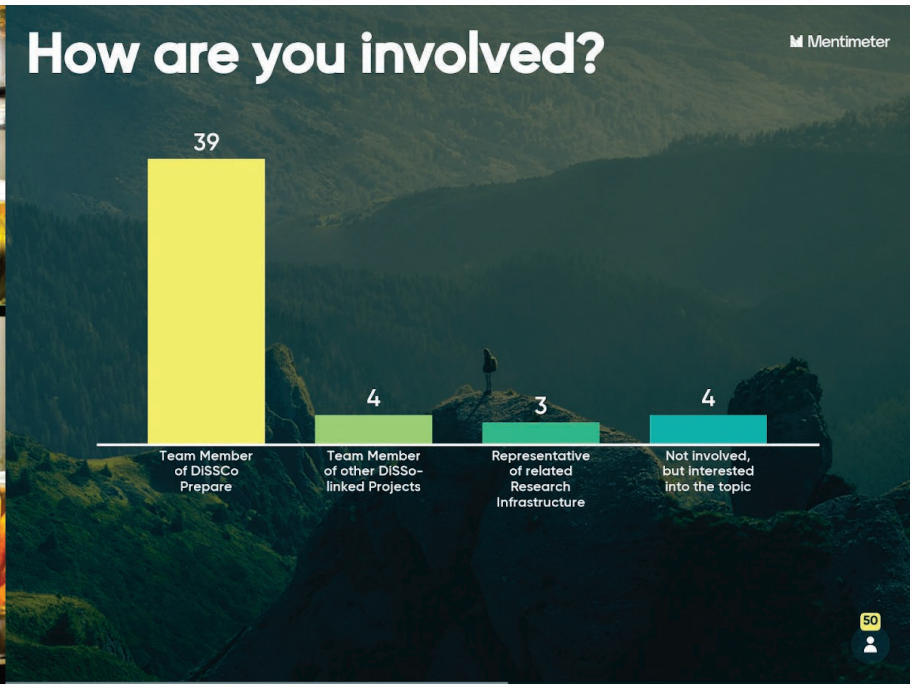
The DiSSCo Knowledge Base - https://dissco.tech/2020/12/18/the-dissco-knowledgebase/

## Participants

More than 50 people participated in the workshop. Mentimeter surveys were used for peoples' introduction and to collect expectations.

## Introduction of participants using Mentimeter


What is your your job title?
programmer
officer
architect
executive
herbarium
data standards
data hero
biodiversity data scntst
unit director
assistant secretary gener
biodiversity informatics
trainer
project
research scientist
cio
professor
communication
data architect
systematics
curator
project manager
head of department
bdinformatics
project coordinator
researcher
liaison officer
junior professor
problem solver
head of division
business analyst
senior project manager
assistant professor
software developer
data scientist
node manager
lecturer
science programme lead
senior research scientist
software engineer
manager
communications manager
science policy coordinato
science
collections manager
project management
research associate
data
scientific
director
LIKE A BOSS
39

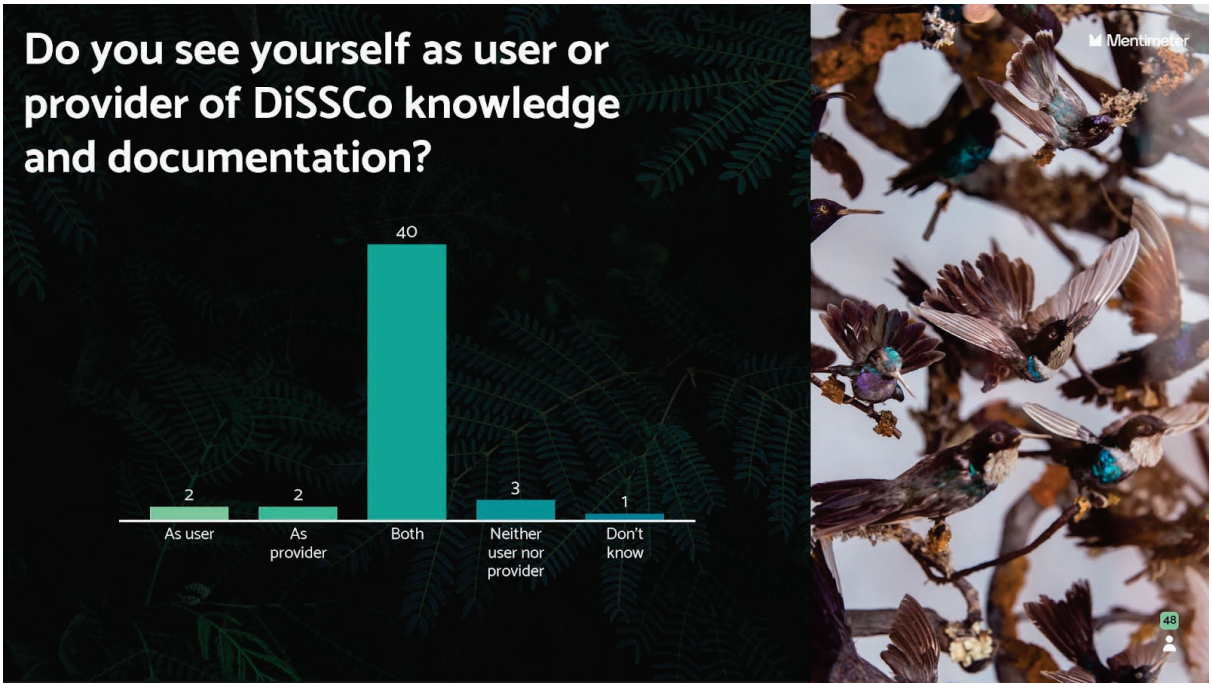

How are you involved?
Mentimeter
39
4
3
4
Team Member of DISSCo Prepare
Team Member of other DISSo-linked Projects
Representative of related Research Infrastructure
Not involved, but interested into the topic
50


## Introduction to the Knowledgebase

The presentation introducing the DiSSCo Knowledgebase (beta version) is available at: https://dissco.teamwork.com/#/files/9973261

## Identification of stakeholder groups

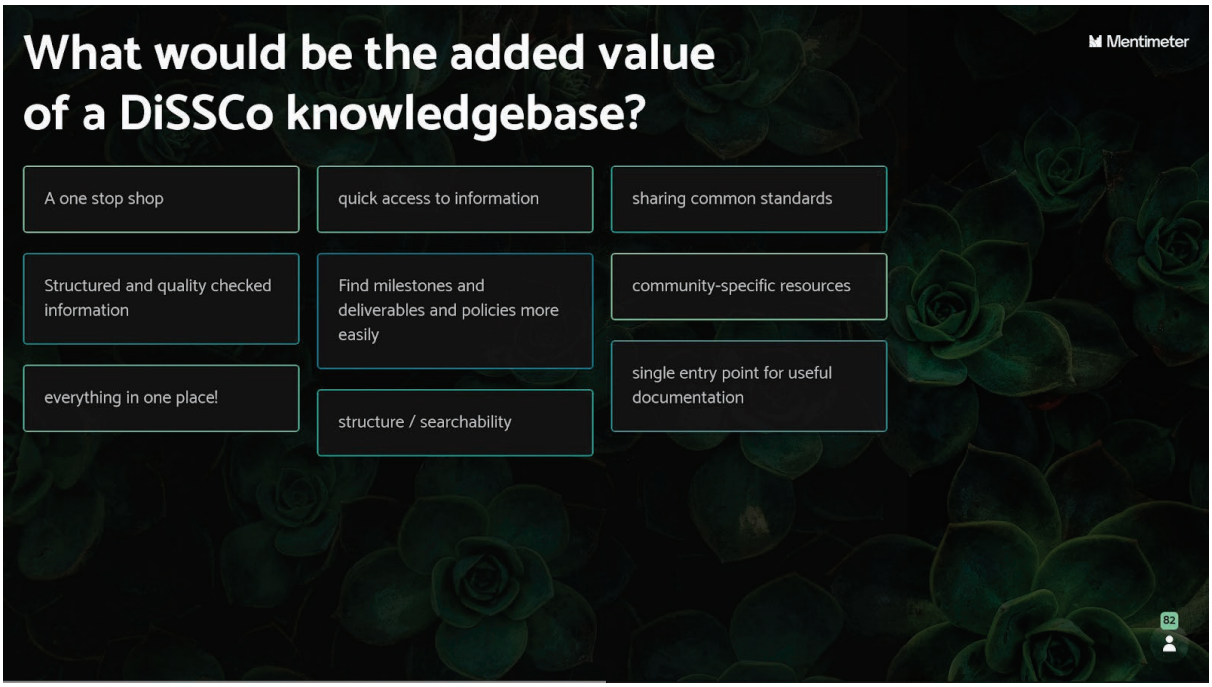
Mentimeter surveys were used to get feedback on existing knowledge sources as well as on the added value of the DiSSCo Knowledgebase.

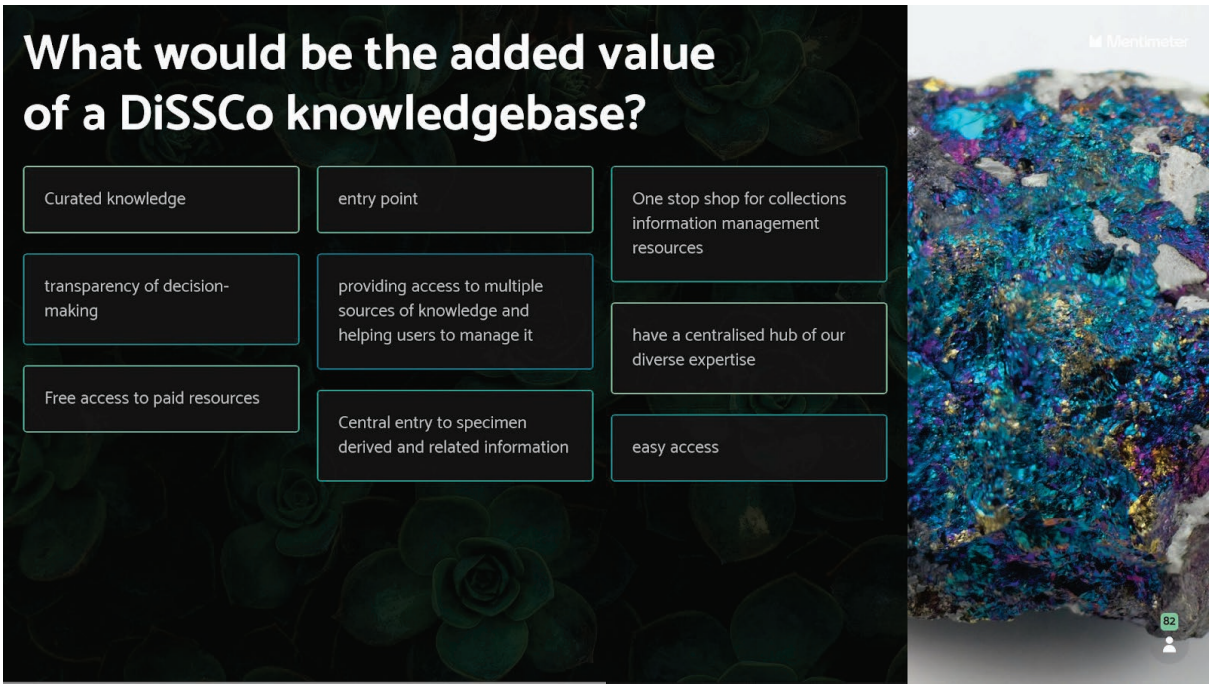


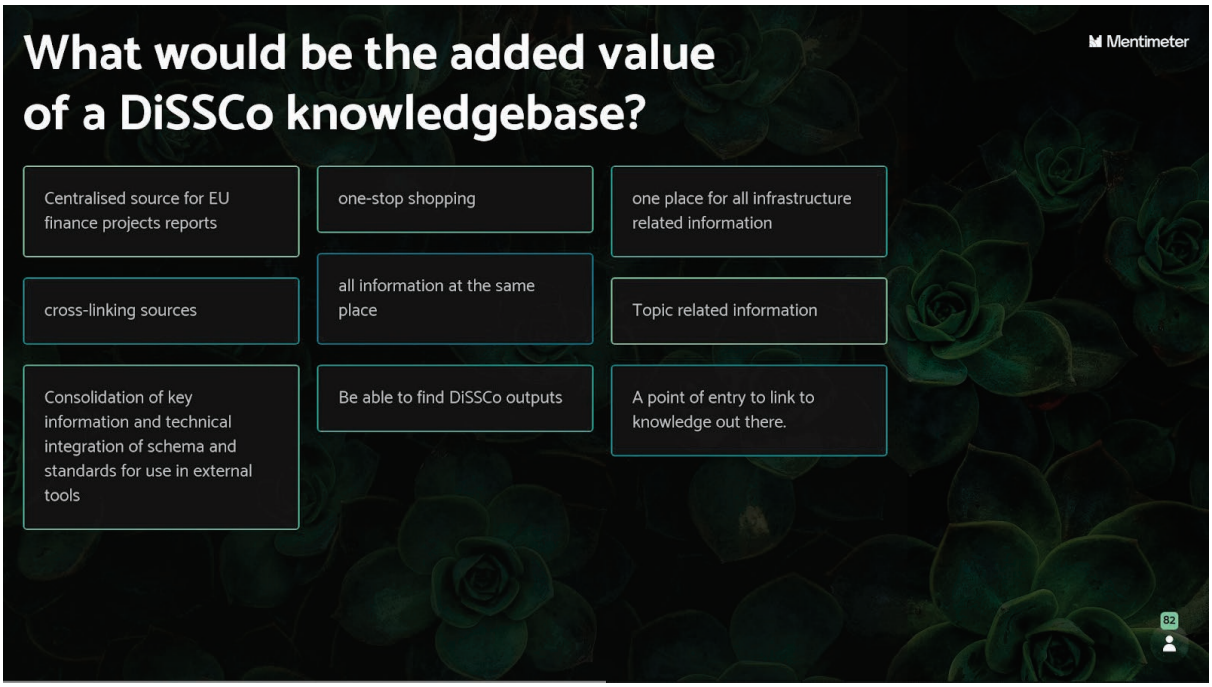
## What are the main sources of knowledge and documentation you use in your daily work?

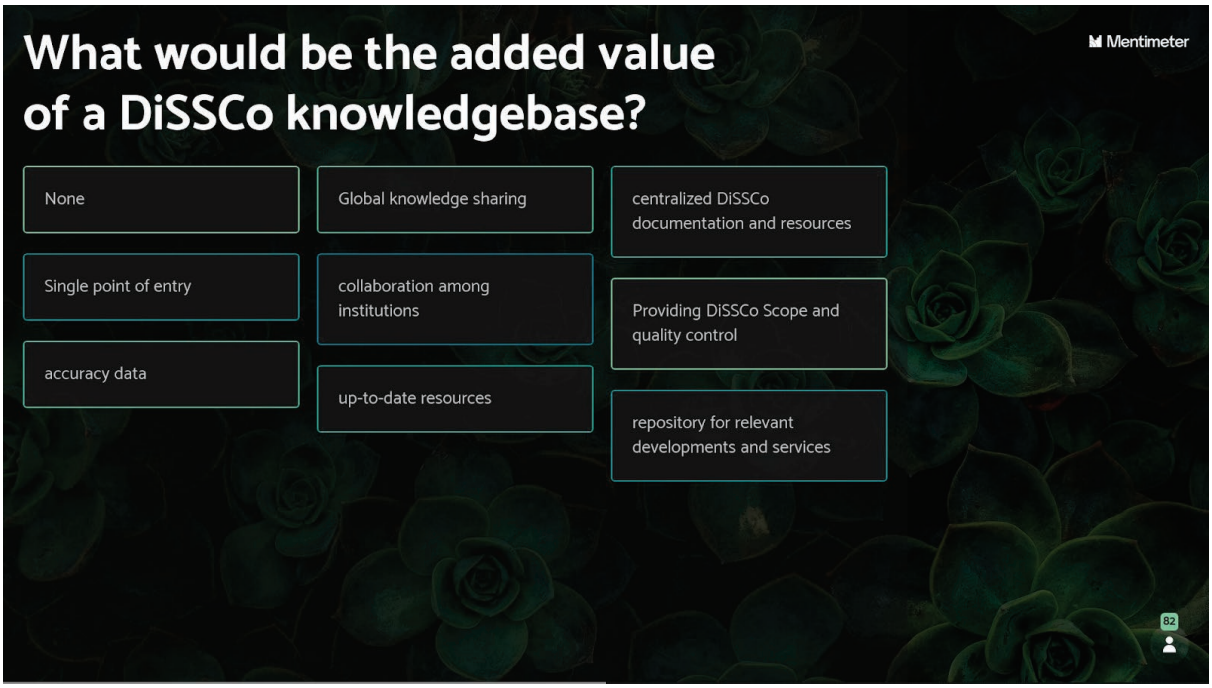


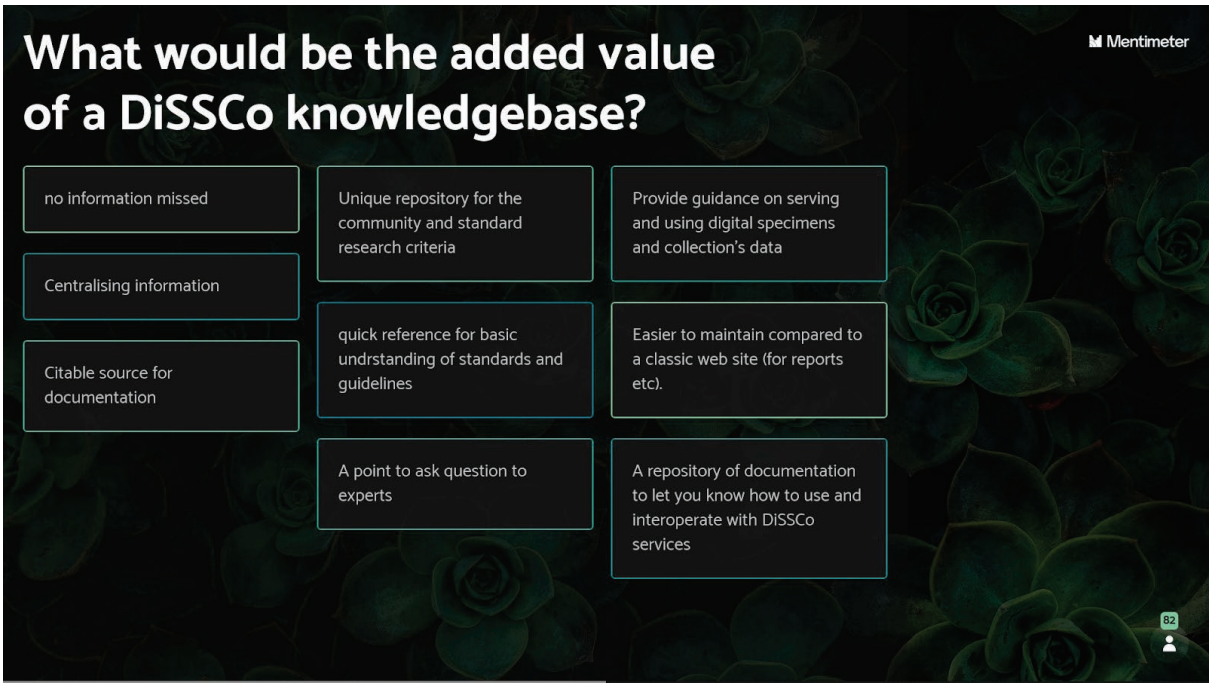
## What would be the added value of a DiSSCo knowledgebase?

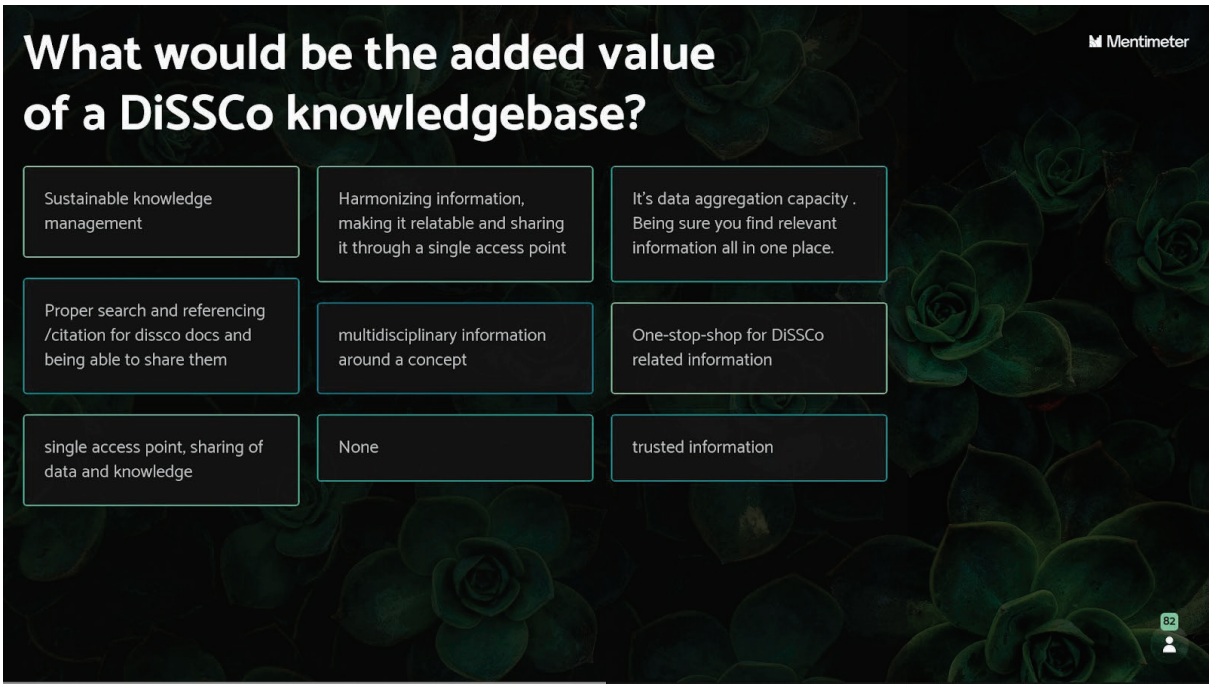

What would be the added value of a DiSSCo knowledgebase?
Mentimeter
A one stop shop
quick access to information
sharing common standards
Structured and quality checked information
Find milestones and deliverables and policies more easily
community-specific resources
single entry point for useful documentation
everything in one place!
structure / searchability
82

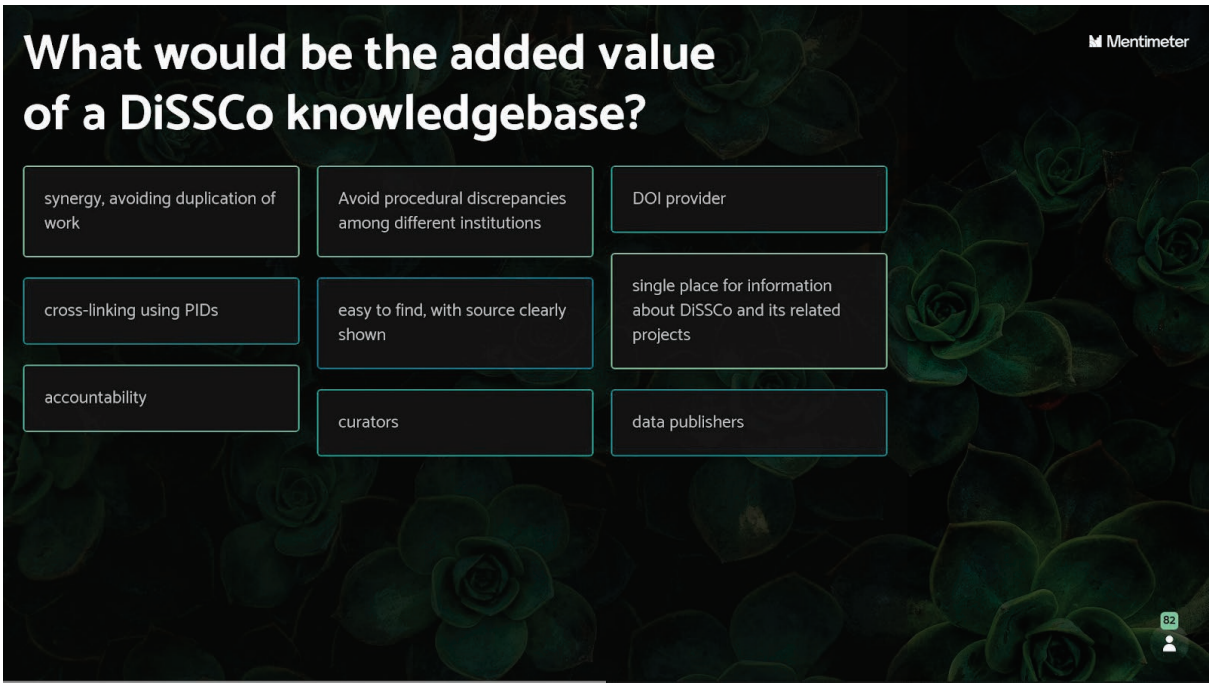

What would be the added value of a DiSSCo knowledgebase?
Curated knowledge
entry point
One stop shop for collections information management resources
transparency of decision-making
providing access to multiple sources of knowledge and helping users to manage it
have a centralised hub of our diverse expertise
Free access to paid resources
Central entry to specimen derived and related information
easy access
82

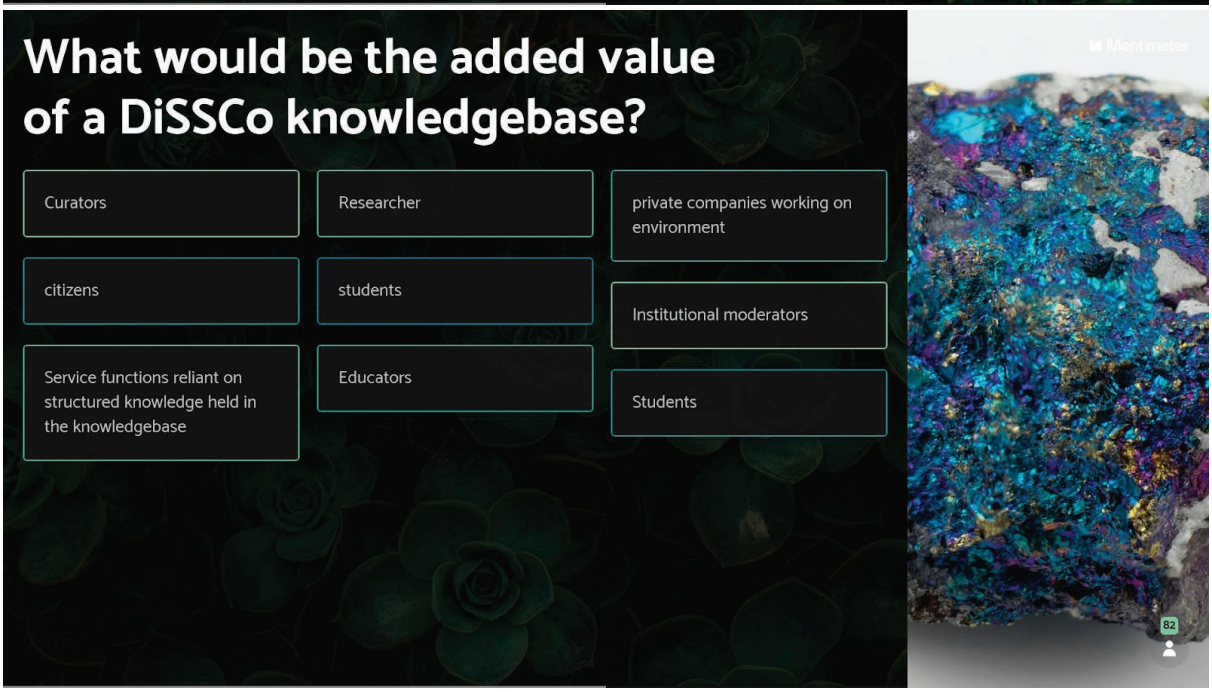

What would be the added value of a DiSSCo knowledgebase?
Mentimeter
Centralised source for EU finance projects reports
one-stop shopping
one place for all infrastructure related information
cross-linking sources
all information at the same place
Topic related information
Consolidation of key information and technical integration of schema and standards for use in external tools
Be able to find DiSSCo outputs
A point of entry to link to knowledge out there.
82

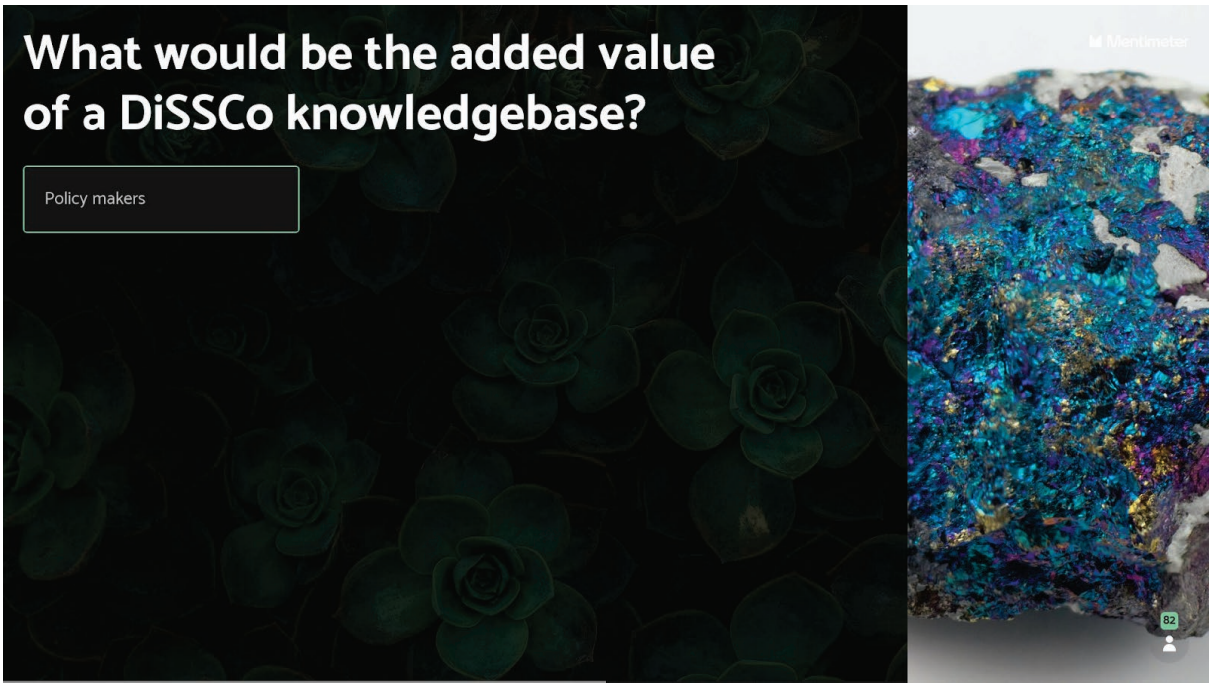

What would be the added value of a DiSSCo knowledgebase?
Mentimeter
None
Global knowledge sharing
centralized DiSSCo documentation and resources
Single point of entry
collaboration among institutions
Providing DiSSCo Scope and quality control
accuracy data
up-to-date resources
repository for relevant developments and services
82


What would be the added value of a DiSSCo knowledgebase?
Mentimeter
no information missed
Unique repository for the community and standard research criteria
Provide guidance on serving and using digital specimens and collection's data
Centralising information
quick reference for basic undrstanding of standards and guidelines
Easier to maintain compared to a classic web site (for reports etc).
Citable source for documentation
A point to ask question to experts
A repository of documentation to let you know how to use and interoperate with DiSSCo services
82

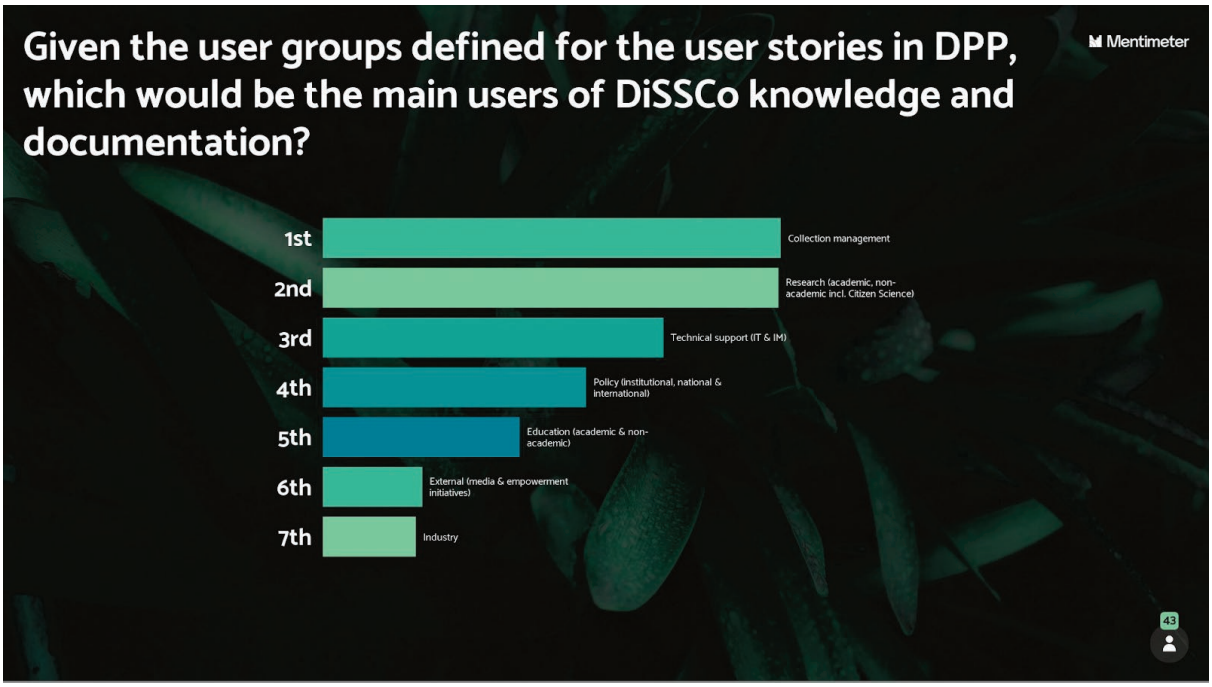

What would be the added value of a DiSSCo knowledgebase?
Mentimeter
Sustainable knowledge management
Harmonizing information, making it relatable and sharing it through a single access point
It's data aggregation capacity . Being sure you find relevant information all in one place.
Proper search and referencing /citation for dissco docs and being able to share them
multidisciplinary information around a concept
One-stop-shop for DiSSCo related information
single access point, sharing of data and knowledge
None
trusted information
82


What would be the added value of a DiSSCo knowledgebase?
Mentimeter
synergy, avoiding duplication of work
Avoid procedural discrepancies among different institutions
DOI provider
cross-linking using PIDs
easy to find, with source clearly shown
single place for information about DiSSCo and its related projects
accountability
curators
data publishers
82


What would be the added value of a DiSSCo knowledgebase?
Curators
Researcher
private companies working on environment
citizens
students
Institutional moderators
Service functions reliant on structured knowledge held in the knowledgebase
Educators
Students
Mentimeter
82




The final two questions were related to stakeholders of the knowledgebase. Which groups would need access to the knowledge and which DiSSCo User Groups (compare DPP WP1) would be main users of the DiSSCo Knowledgebase?




After the mentimeter surveys the stakeholders were discussed further and whether also internal users are included in the different groups identified.

Based on the Mentimeter results and discussion it was decided to focus on 5 Stakeholder Groups:

- Collection staff (curators, collections managers & data managers, etc.)
- Researchers & students
- Developers
- Policy makers & funders
- Citizens, citizen scientists & wider public

## How to best organize the knowledge for different stakeholders?

- It needs to be clear: What are the boundaries of the knowledgebase?
- The Knowledgebase should behind the scene connect the different e-Services
- To improve organization, collection of questions and usage of semantics to group information to fit the needs of users
- Should the KB be responsive to user needs? Respond to questions or is this all covered by the Helpdesk?
- How should we link other e-Services?

- Would a Community Forum be an added value?
- Possible need for providing (to a certain extent) different access paths to the KB for different user profiles (e.g. the non-initiated, the insiders, etc.)
- External Users: First information hub could be the Website not the KB
- Are internal users - project members the biggest user group?

## Places where to make knowledge available?

- DSpace implementation, what is needed in addition?
    - GitHub
    - Training systems/delivery
- DiSSCo website and comms (Binnacle Blog? Social etc.)
- How can we reach different stakeholder groups?
- We do not need support for collaborative work on documentation, but a stable reference point
- KB could not only be a repository, but a trusted resource for the community, guidance on how to do things, need for community-resource they can rely on.

## AOB

- Handbook on what kind of formats / information could be made available in DSPACE, e.g. HTML/XML/MD and PDF?
- Storage of the editable versions? At the moment using PDFs means we lose access to assets (data/images/figures).
- Stakeholder Group Publisher: More Information on the Publishing functionality needed, so we can publish more sustainable and reusable publications, have project assets be a bit more FAIR - including presentation assets for DiSSCo in general.

## Breakout Rooms

### Task

Identify documentation and outputs for identified stakeholders that need to go into the knowledge base (both existing and not yet existing documentation)

*If you discuss, please keep in mind the information types identified and presented earlier as well as systems which could be a home for the kind of knowledge collected for the Stakeholder group discussed.*

#### Breakout Group 1 - Curators, collection & data managers

**Summary:**

- overview page or “directory” of most relevant services and checklists
- best practices and guidelines for special collections, incl. sampling practices
- workflow descriptions, connecting different workflows within the collections
- use of standards, mapping (collections descriptions), link between physical and digital specimens
- provide workflows, best practices and training materials in different languages (start English but aim for multilingual content)
- links to training materials / policy documents incl. legislation

- link to self-assessment tool on Digital maturity of institutions:
    - provide information about expertise and strengths of certain institutions
    - allow for synergies, e.g. working groups for collections in similar situation, facing similar challenges (exchange of information and expertise)

### Breakout Group 2 - researchers, students, teachers

| **What documents?** | **In what form?** | **Where?** | **Who?** |
|---|---|---|---|
| Public Documents & External Resources (= Literature) | A general repository, different access options and filters related to the user/stakeholder | | |
| What is DiSSCo? What can DiSSCo provide for me? "Where to start?" documentation (summarized, accessible language information) | | Google it? | every first time user of DiSSCo |
| Use cases | Visuals, static documents<br><br>from researcher point of view: no need to be very interactive | | |
| Training material | interactive material | | |
| Software | | | |

General comment from the breakout group: especially for the researchers point of view, the use cases need to be very visually present. Most researchers are asking the question: how can I benefit from DiSSCo?

### Breakout Group 3 - Developers

**General points:**

- no code should be stored directly in the KB
- proper metadata should be associated with repository links so that the repositories are findable and understandable (what does the repository do/provide?)
- GitHub would work as an external resource to the KB
- could potentially pull metadata from GitHub automatically
- KB should store guidelines for both internal and external developers
- the metadata and developer guidelines need to be curated
- KB should serve as a developers interface between DiSSCo, GBIF, ALA, iDigBio, COL, etc.

Breakout Group 4 -Policy makers, funders

Policy makers can mean those who make or contribute to internal policies and governments/civil servants/external wider policy makers. Funders could be government (national or international); quasi-governmental bodies such as research funding bodies; or companies, private individual donors etc with a wide range of different needs - generally though they will need some form of information about the costs and impact of what they might fund.
Within this group we have no-one who self-identifies to this user group but many who have some relevant experience - with internal or external policy makers and bidding to funders. DiSSCo Prepare tasks on **socio-economic indicators** will be very relevant to these stakeholders.

We need to be able to find information to convey to policymakers and funders at the right time? They are unlikely to look for it?
They want to the point summaries / answers to questions.
Could consider new formats of information such as **policy briefs** on key topics? What would the scope of these be?

## Discussion on the scope of policy requirements for the knowledgebase

Five major topics emerged:

**Performance indicators** for DiSSCo are very relevant to funders and policymakers. We want reports or dashboards covering the breadth of DiSSCo activities. They need to answer on progress to the national governments, but also to the EU or higher international instances, where they need to collect progress of RIs among others.

This is associated with providing support with evidence to the RI evaluation exercises by ESFRI, which guidelines and criteria are available at the Public Guide.

Our **institutional policies** are a key document set for internal policy makers, and for DiSSCo centre. We need the knowledge base to hold these documents, or at least metadata about them, to help assure developing policy alignment between DiSSCo institutions (this will underpin the functionality of the Task 7.3 policy tool), i.e. holds the metadata schema, which could be used to tag/mark up policy documents to help them be more discoverable and accessible.

Our **policy standards and policy metadata schema** are central to a classification of our policies and provide a route to understand their alignment across different institutions, and with the DiSSCo Policies.

The **DiSSCo policies** associated with the delivery and needs of DiSSCo Services and the requirements of DiSSCo to use these services. Ideally these need to be described in the same way as the Institutional Policies (e.g. via the same metadata schema) to enable us to align institutional and DiSSCo policies.

**National, European and International Policies** relating to biodiversity (e.g. IPBES Indicators) addressing larger societal challenges associated with our domain.

### Breakout Group 5: Other external stakeholders: citizens, research infrastructures, industry

Documentation and Outputs

**Identify relevant content for the (external) stakeholders**

- Maybe distinguish a bit more the "citizens" according to use cases
- we won't know all the (future) use cases and stakeholders, thus we need to react flexibly to user needs.
    - compiled information on services and communities that could help / support with certain use cases (e.g. analytical data)
    - How to curate sets of documents? (e.g. a shopping basket in the KB)
    - Set of questions to recommend documents to users according to stakeholder groups
- FAQ & glossary
- DiSSCo website vs. DiSSCo KB
    - website could deep link to KB
    - latest information on the progress (e.g. progress reports) could live in the KB

- What to do with data that is not in the primary scope of DiSSCo?
    - How to make available our data?
    - How to link?

**Increase Discoverability and Relevance for (external) stakeholders**

- DiSSCo RI could feed back cross-linked literature etc. to the KB
    - maybe ingest results of literature/resource mining in the KB
- based on usage statistics suggest content (e.g. "Other users also searched for …", "This might be interesting to you as well…") based on the queries the users do.

- enrich KB with SEO metadata in order to make knowledge findable via Google and other search engines (people will search these instead of directly on DiSSCo KB)

- for citizens: tease / highlight interesting content (stories), specimens etc.
    - linking infrastructures and storytelling could be a by-product from BiCIKLE project
    - How to curate private collections (held by citizens)? How to digitize them and contribute?

- put together existing documents and then identify gaps in information types
    - results from iDigBio
    - and others ...
- include infrastructure studies (socio-technological aspects of data infrastructures) as a resource

## Identifying next steps and closing

- Share content which should be available in the KB
- Suggestion: Board of editors after the project time for new content and curation