

# DiSSCo related output

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

## Title

The DiSSCo Knowledgebase

## Author(s)

Petersen, Mareike  
von Mering, Sabine  
Glöckler, Falko  
Pim Reis, Julia

## Identifier of the author(s)

<https://orcid.org/0000-0001-8666-1931> (MP)  
<https://orcid.org/0000-0003-2982-7792> (SvM)  
<https://orcid.org/0000-0002-7127-2738> (FG)  
<https://orcid.org/0000-0002-5357-6148> (JR)

## Affiliation

Museum für Naturkunde – Leibniz Institute for  
Evolution and Biodiversity Science  
(<https://ror.org/052d1a351>)

## Publisher

## Identifier of the publisher

## Resource ID

## Publication year

2020

## Related identifiers

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

Yes

## Creation date

18/012/2020

## Version

1

## Abstract

As an initiative formed by public research institutions, DiSSCo is committed to Open Science. We believe that Open Science not only makes the 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 are openly available as much as possible and research data easily Findable, more Accessible, Interoperable and Reusable (FAIR principle).

DiSSCo Prepare (DPP), the preparatory project phase of DiSSCo, will build on profound technical knowledge from various sources and initiatives. In order to allow for efficient knowledge and technology transfer for partners building the DiSSCo technical backbone, a central and freely accessible DiSSCo Knowledgebase will be designed and implemented within the project. The conceptual and developmental work is done under the Work Package “Common Resources and Standards” and the Task “DiSSCo Knowledgebase for technical development” (both led by the Museum für Naturkunde Berlin). This hub for knowledge management relevant within the DiSSCo context will not only store all research outputs from DiSSCo-linked projects in one place but also act as a reference for further building blocks relevant for the DiSSCo Research Infrastructure (RI).

## Content keywords

scientific

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

**WP number**  
WP5

**Project output**  
Techno Blog

**Deliverable/milestone number**  
MS5.5

**Dissemination level**  
Public

**Rights**

**License**  
CC0 1.0 Universal (CC0 1.0)

**Resource type**  
Text

**Format**  
PDF

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

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



## DiSSCo Prepare

**WP Lead:** Mareike Petersen (MfN)  
**Task Lead:** Mareike Petersen & Falko Glöckler (both MfN)  
**Authors:** Mareike Petersen, Sabine von Mering, Falko Glöckler, Julia Pim Reis (all MfN)

---

## Abstract

As an initiative formed by public research institutions, DiSSCo is committed to Open Science. We believe that Open Science not only makes the 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 are openly available as much as possible and research data easily Findable, more Accessible, Interoperable and Reusable (FAIR principle).

DiSSCo Prepare (DPP), the preparatory project phase of DiSSCo, will build on profound technical knowledge from various sources and initiatives. In order to allow for efficient knowledge and technology transfer for partners building the DiSSCo technical backbone, a central and freely accessible DiSSCo Knowledgebase will be designed and implemented within the project. The conceptual and developmental work is done under the Work Package “Common Resources and Standards” and the Task “DiSSCo Knowledgebase for technical development” (both led by the Museum für Naturkunde Berlin). This hub for knowledge management relevant within the DiSSCo context will not only store all research outputs from DiSSCo-linked projects in one place but also act as a reference for further building blocks relevant for the DiSSCo Research Infrastructure (RI).



H2020-INFRADEV-2019-2  
Grant Agreement No 871043

## The DiSSCo Knowledgebase

Authors: Mareike Petersen\*, Julia Pim Reis\*, Sabine von Mering\*, Falko Glöckler\*

\* Museum für Naturkunde Berlin, Germany

### Introduction

As an initiative formed by public research institutions, DiSSCo is committed to Open Science. We believe that Open Science not only makes the 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 are openly available as much as possible and research data easily Findable, more Accessible, Interoperable and Reusable ([FAIR principle](#)).

[DiSSCo Prepare](#) (DPP), the preparatory project phase of DiSSCo, will build on profound technical knowledge from various sources and initiatives. In order to allow for efficient knowledge and technology transfer for partners building the DiSSCo technical backbone, a central and freely accessible DiSSCo Knowledgebase will be designed and implemented within the project. The conceptual and developmental work is done under the Work Package “Common Resources and Standards” and the Task “DiSSCo Knowledgebase for technical development” (both led by the Museum für Naturkunde Berlin). This hub for knowledge management relevant within the DiSSCo context will not only store all research outputs from DiSSCo-linked projects in one place but also act as a reference for further building blocks relevant for the DiSSCo Research Infrastructure (RI).

### Approach

As a first step, the extent of information types expected to be stored in the knowledgebase was collected. To get the most complete picture we discussed this topic within the respective project task group and work package, but also together with project overarching bodies such as the DiSSCo Technical Team. As a last preparatory step, we sent a [survey](#) 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 planning steps. The latest overview of desired information types is given in Figure 1.



**Figure 1:** Information Types in the DiSSCo Knowledgebase. Expected cluster of information categories (blue dots) based on DPP Project outcomes and relevant external resources. The format of resources varies within and among information types.

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.

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 information type “Public Documents and External Resources” in order to aggregate references to distributed documents and sources in a single point of entry. A comprehensive landscape analysis with short presentations of each system took place during two task group meetings. 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](#).

### First Version

The implementation of DSpace as a first version of the DiSSCo Knowledgebase core will have a customized layout with the DiSSCo branding. It will allow to create a hierarchy of DiSSCo-linked projects and their respective collections of documents and references. In order to store content like Frequently Asked Questions (FAQs), best practices, guidelines, recommendations and documented decisions on the RI, the DiSSCo partners will be enabled to extend the knowledgebase with their content (being free text or files) with the help of easy-to-use web-forms that include a rich text editor. An editorial workflow modelled in the system will allow the platform administrators to review the content prior to publication via role based access. This will also allow for preparing documents privately before publishing them and

conducting a profound quality assurance.

The first version of the DiSSCo Knowledgebase will be launched by end of January 2021 at <http://know.dissco.eu>

### Next Steps

As a next step, the current results of the implementation of the DiSSCo Knowledgebase will be presented at the first [All Hands \(virtual\) Meeting](#) of DiSSCo Prepare (18 – 22 January 2021). This is an event that will bring together leaders and partners of the project, with the objective to present, discuss and produce key elements of what will become Europe's leading natural science collections Research Infrastructure, DiSSCo RI. In a dedicated session, the participants will have the opportunity to test the first version of the knowledgebase by browsing the software and testing the features, allowing us to collect feedback and requirements from the project partners.

The DiSSCo Knowledgebase, in its final version, will provide 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; an assessment of their technical readiness for DiSSCo as well as specifications on their relevance for the overall DiSSCo technical infrastructure and the DiSSCo data model.

### Outlook

DiSSCo uses a DOI namespace provided by [DataCite](#) for assigning DOIs to documents like public deliverables and reports. This process will be automated with the help of a DSpace plugin on the document's submission. In addition, depositing and linking documents on [Zenodo](#) will be integrated.

To increase the findability of content the metadata will be linked and enriched by cross-references to related content and external resources (e.g. ORCID). In order to optimize the findability even outside the search interface of the knowledgebase the JSON-LD format will be embedded in the landing pages, so the visibility of DiSSCo outcome and knowledge is maximized in the big search engines.

Over the course of the upcoming year 2021 all the other information types will be accommodated or linked in the DiSSCo Knowledgebase. This can be assured by submitting at least a metadata description about information that will be managed outside of DSpace (e.g. software code on [GitHub](#) or controlled vocabulary in [WikiBase](#)). But by providing machine-readable formats, custom plugins in DSpace will allow even richer connections between different components of the DiSSCo Knowledgebase.

Want to get involved? Feel free to check our remote repository on [GitHub](#) or contact us [here!](#)