SYNTHESYS+

Synthesis of Systematic Resources

a DiSSCo project

D2.1 Consolidated document on policy topics, related best practices and standards

Laura Tilley[1], Matt Woodburn[2], Ana Casino[1], Alan Paton[3], Edmund Schiller[4], Karin Wiltschke[4], Lisa French[2], Frederik Berger[5], Laurence Bénichou[6], Gabor Csorba[7], Jonathan Blettery[6], Elspeth Haston[10], Patricia Mergen[11], Carole Paleco[9], Francisco Pando[8], Patrick Semal[9], Vince Smith[2]

[1]CETAF, [2]NHM London, [3]Royal Botanic Gardens Kew, [4]Natural History Museum Wien, [5]Museum für Naturkunde, [6]National Museum of Natural History, [7]Hungarian Natural History Museum, [8]Real Jardín Botánico, [9]Royal Belgian Institute of Natural Sciences, [10]Royal Botanic Garden Edinburgh, [11]Meise Botanic Garden

Grant Agreement Number | 823827

Acronym | SYNTHESYS PLUS

Call | H2020-INFRAIA-2018-2020

Start date | 01/02/2019

Duration | 48 months

Work Package | [WP2 D2.1 Coordinate harmonisation of policies and facilitate the implementation of policy mandates]

Work Package Lead | [Ana Casino, Laura Tilley]

Delivery date | [28.09,2022]

Summary

Task 2.1 has delivered a policy metadata schema. A metadata schema is a system/model used for labelling, tagging, recording and cataloguing information in a structured way via defined controlled vocabularies, and rules (e.g. which data formats to use, how data relates to each other). In this case, the metadata schema is designed to catalogue and centralise the policies that DiSSCo services need institutions to align with, and data on institution's existing policies. The metadata schema forms the backbone of the DiSSCo Policy Self-Assessment tool that was developed in DiSSCo Prepare Task 7.3. Forming the backend of the tool the metadata schema provides a structure that allows institutions to map their existing policies against DiSSCo service needs, and makes the policy data more interoperable and comparable for analyses and producing data visuals within the tool. It was decided to focus on using only one of the proposed DiSSCo services, the European Loans and Visits System (ELViS) as a case study for building and testing the metadata schema. The reason being that ELViS is the most significantly developed service thus allowing for a clearer understanding of what policies may be required for its areas of functionality. It also covers an array of policy topics that are likely to be relevant to other DiSSCo services. Development of the schema was done in 5 phases that included the identification and defining of ELViS policy needs; conceptualisation of the metadata schema data model for the DiSSCo Self-Assessment Policy tool and incorporation of the identified policies relevant to ELViS within it; presenting a first draft of the schema to a wider audience for feedback (Milestone 24 workshop); and finalisation of the data model and incorporation of the policies. Care was taken to ensure that the schema is flexible to eventually incorporate other DiSSCo services. The relevant ELViS policies that institutions need to align with were defined based on evidence found in existing DiSSCo documentation, and consulting with ELViS coordinators. The finalised policy metadata schema is presented within this deliverable report (section 2.2) in which is presented the major categories that structure the envisaged relevant policies and the ELViS service, and institutional details. Policies have been described and categorised using an amended form of the hierarchical category system used in the ICEDIG 7.1 policy survey. In this categorisation, ‘Policy Category’ is the highest level in which a policy can be grouped, followed by ‘Policy Area’ and finally ‘Policy Component’ describing a specific element within a ‘Policy Area’. Altogether there are 2 ‘Policy Categories’, 9 ‘Policy Areas’ and 40 Policy Components. ELViS has been broken down into 10 ‘Service Components’ that refer to specific aspects of the service workflow (e.g. Providing People Data; Providing Specimen Data).

As part of this deliverable a case study of mapping the T2.1 institutions policy data to the metadata schema was conducted in order to test the practicality of the metadata schema and to provide a preliminary analysis of the current state of natural sciences institutional policies alignment towards ELViS policy needs. The major results reveal that institutions could map their policy information to the metadata schema. Alignment is generally good for Policy Areas that contain components that relate to EU level legal requirements such as GDPR, Access and Benefit sharing; but less good for Policy Areas with Components that refer to concepts that are new or still under development such as the use of the standard Minimum Information for A Digitised Specimen (MIDS), and the TDWG Collection Description standard. In some cases institutions followed policies, but as best practice or informal procedures rather than having formal documentation in place, for example in some aspects of data publication, and data management.

The concluding remarks (section 4) include recommendations for improvements to the current metadata schema, how DiSSCo can facilitate institutional alignment, future expansion, sustainability and maintenance.

Key Words: DiSSCo RI, ELViS, policies, natural science collections, metadata schema.

1. Introduction

1.1 Context

The full realisation of the Distributed System of Scientific Collections Research Infrastructure (DiSSCo RI), with its aim of unifying European natural assets under common curation and access to make data easily findable, accessible, interoperable and reusable (FAIR); will require a certain level of policy alignment from its distributed network of European Natural Science Institutions (NSI). Although it is not DiSSCo's intention to interfere with the internal operations of institutions and their management of collections, some institutional policy alignment is needed to allow for the seamless flow of FAIR data, for the creation of a detailed common research agenda for the DiSSCo consortium, and for the infrastructure to provide distributed services at a consistent level of optimal quality. This is a large undertaking and it will require the establishment of sustainable mechanisms and tools to streamline the communication of the

policy alignment needs and their purpose to DiSSCo partners, as well as to provide support and guidance for institutions for assessing their current policy status and progress towards alignment.

Task 2.1 (T2.1) '*Coordinate harmonisation of policies and facilitate the implementation of policy mandates'* under the SYNTHESYS+ project is led by the Consortium of European Taxonomic Facilities (CETAF) and involves 9 partner institutions,

- Hungarian National History Museum - Budapest (HNHN),
- Meise Botanic Garden (BGM),
- Natural History Museum Wien - Vienna, (NHMW),
- Royal Botanical Garden Kew (RBGK),
- Royal Botanical Garden Edinburgh (RBGE), Real Jardin Botanico - Madrid (CSIC),
- Museum für Naturkunde Berlin (MfN),
- National Museum of Natural History - Paris (MNHN), and
- Royal Belgian Institute of Natural Sciences - Brussels (RBINS).

The initial objective of the task was to collate and harmonise policies and best practices already adopted by institutions (mainly focused on partners of the task). In the early stages a survey developed within Task 7.1 of the 'Innovation and consolidation for large scale digitisation of natural heritage' (ICEDIG) project about their existing policies related to Natural Science Collections (NSCs) was sent to SYNTHESYS+ T2.1 partner institutions. The policy information received was then analysed for gaps, commonalities and areas requiring harmonisation. ICEDIG Task 7.1 identified policies relevant to NSCs, collected policy information from six institutions and, as a final deliverable, created a policy dashboard showing the current status of policies, e.g. gaps, commonalities, strengths and weaknesses (Agosti and Egloff, 2019). Given the work undertaken within ICEDIG and the amount of effort required within this Task T2.1, it became apparent that it was not a possible long term approach for DiSSCo to manually collate, analyse and harmonise all its partner's institutional policies. The NSC policy landscape is too heterogeneous, with different policies having different levels of enforcement (e.g. national or institutional level). Policies can also become outdated very quickly and the definition of a policy can be broad, for instance, they may be a formal EU/international requirement, or best practices or guidelines. Task 2.1 decided in collaboration and alignment with DiSSCo Prepare Task 7.3 (Develop and establish DiSSCo policies), led by the Natural History Museum London (NHM), to create instead a more dynamic and sustainable deliverable.

1.2 Scope

The revised aim of T2.1 is to develop a metadata schema, describing different policy elements that may be necessary for institutions wanting to collaborate in DiSSCo services. At the same time T2.1's development process is a case study to test the methodology of building a policy metadata schema for DiSSCo services. A metadata schema is a system/data model used for labelling, tagging, recording and cataloguing information in a structured way via defined controlled vocabularies and rules (e.g. which data formats to use, how data relates to each other). In this case, the metadata schema is designed to structure and centralise all relevant policy elements of DiSSCo services and data on the institution's existing policies. The metadata schema forms the backbone of the DiSSCo Policy Self-Assessment tool which was developed in DiSSCo Prepare WP7, T7.3 (**Figure 2 and 2**). Furthermore, it forms the basis of the questions used by the tool. Through the Policy Self-Assessment tool DiSSCo Partners can map their institutional policies against the envisaged requirements of DiSSCo's ELViS service and assess the status of their policy alignment visually via an institutional summary.

The schema will also link to reference policies, best practices, guidelines of other institutions (if openly accessible) via the DiSSCo Knowledge Base (https://know.dissco.eu/), in order to help with areas of poor alignment. Moreover, via a projected dashboard component the schema will provide the Coordination and Support Office (CSO) and DiSSCo partners with an overall view of the state of policy compliance and gaps across the consortium, thus helping to identify areas where general support is needed (Humphries et al., 2022).

Task 2.1 decided to focus on establishing which policy elements may be alignment requirements for the Service '*European Loans and Visits system'* (ELViS) as a case study to test the methodology of building a policy metadata schema for DiSSCo services. ELViS is currently the only service that is significantly developed, thus allowing for a clearer understanding of what policies may be required for its areas of functionality. ELViS will be a one-stop-shop for providing virtual (digitisation on demand), in-house and loan access to NSC specimens and the data associated with them. It will be a transactional platform, so ELViS will not store data but rather, aggregate data on available collections, specimens and institution facilities from open access repositories (e.g. GBIF, CETAF collection registry, institution repositories). As a DiSSCo core service ELViS will be fully integrated in the research infrastructure's digital architecture, and due to its diverse functionality it will cover a wide array

of policy topics, including data standards, digitisation, data publication, elements of loans and collection management. Some of the policy topics will be relevant for other DiSSCo services and so the metadata schema can be expanded to encompass other DiSSCo services in the future.

Task 2.1 also presents a study which applies the T2.1 partners' institutional policy data to the policy alignment criteria (metadata schema) envisaged for using ELViS. This was carried out to understand any challenges of using the metadata schema and to provide an initial analysis of the institutional policy landscape: highlighting true gaps in policies, as well as to identify policies that can be used as examples for other institutions. Based on the outcomes of all the work undertaken, recommendations are provided for further improvements, expansion and sustainability of the metadata schema, as well as on ways to facilitate the policy alignment of institutions in terms of helping to fill in policy gaps.

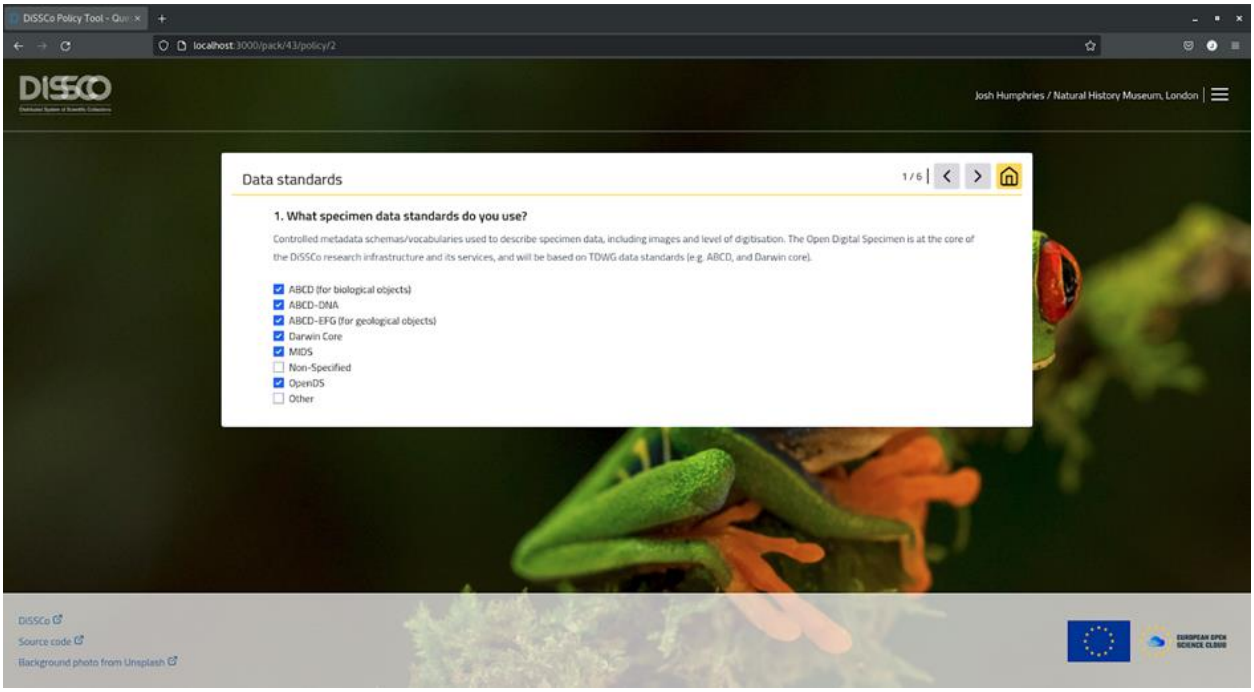


Figure 2. Screenshot of a question from the DiSSCo Policy Self-Assessment tool. Taken from Humphries et al. (2022).

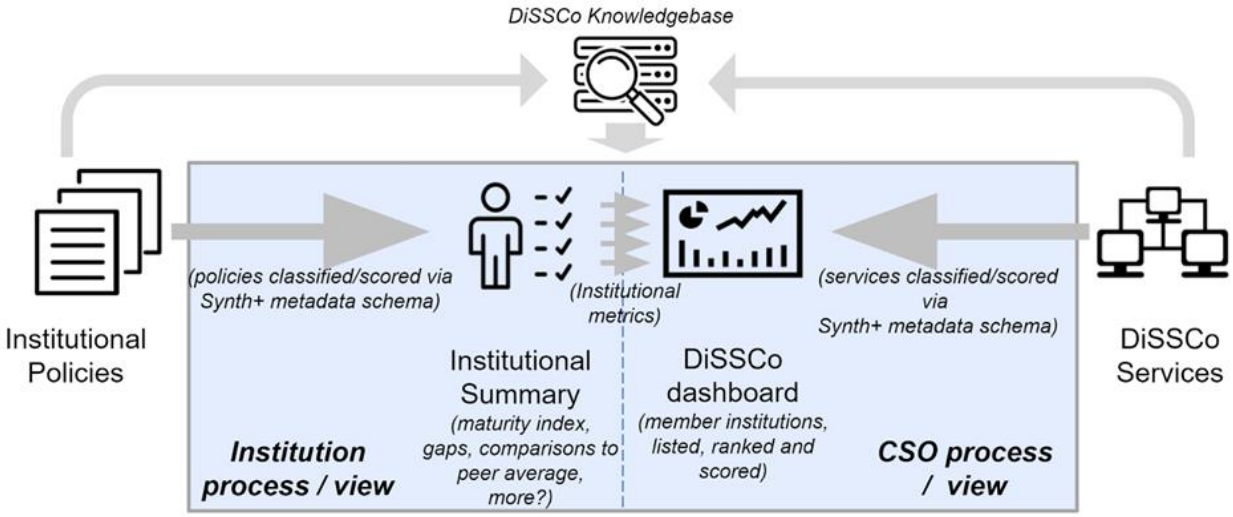


Figure 1. Process diagram for the DiSSCo Policy Self-Assessment tool, showing how the metadata schema is integrated into the tool. Taken from Humphries et al. (2022)

2. Policy metadata schema

This section presents the methodology for developing the metadata schema and the incorporated policy content, and a description of the finalised result for expected ELViS requirements.

2.1 Methodology

The development of the metadata schema was done in five phases between November 2020 and April 2022. The majority of the T2.1 partners were involved in the development, which was coordinated by CETAF via regular online meetings, collaborative tools such as google docs, and Teamwork version 11.3.1 (Teamwork, 2015) . The different phases of development are described in more detail below:

Phase 1 Initial identification of Policy requirements for using ELViS

November 2020 - February 2021

The starting point of developing the metadata schema was to identify the policy content, i.e. the expected policy requirements for using ELViS. This phase helped scope the main structure of the metadata schema. ELViS was first broken down into separate functional components (i.e. Transnational Access to collections, Virtual Access to collection specimens, Loans, Collection discovery, People profile integration, Institution integration etc.) in order to have a harmonised understanding amongst task partners of what ELViS as a service will comprise and how it will operate. This also provided a structure for systematically identifying possible policy requirements for using ELViS, a structure that would minimise the risks of overlooking any aspects. Initially the identification and defining of ELViS components was assisted by the ELViS design document (Islam and Addink, 2019a) and ELViS project plan (Leeuwen van et al., 2019), although not all the components were clearly defined and some interpolation had to be made based on the evidence provided in the documents and the experiences with collection work the Task 2.1 partners provided in regular online meetings.

For each ELViS component, Task 2.1 partners were asked to refer to the policies defined in the ICEDIG 7.1 policy survey and indicate which policies were relevant. Participants were encouraged to provide their responses based on evidence, using what is already known from existing DiSSCo project documentation (e.g. ICEDIG, SYNTHESYS+, DiSSCo Prepare), existing workflows in their own institutions and work being done in other work packages, rather than speculation based on the policies they think are relevant for ELViS, or that institutions need to align with. Interpolation of future policy requirements were considered if there was enough

evidence. To facilitate this work, CETAF created a reference list of existing relevant documentation, and set up bi-weekly to weekly meetings to discuss and agree on the policy needs.

From Phase 1 it became apparent that:

- Not all the policies outlined in the ICEDIG 7.1 survey were relevant for ELViS.
- The general hierarchical structure of policies defined in the ICEDIG 7.1 survey was useful for the policy metadata schema.
- The policy terminology from the ICEDIG survey needed to be more specifically defined and controlled by descriptions/definitions to allow a common degree of understanding of which policy level was required for institutions to align with for using ELViS and why.
- The breakdown of ELViS into its main functional components would be useful for the policy tool for helping institutions to see the relation of why policies were actually needed for the use of the service/what aspect do they actually affect.

Phase 2 Data model and 1st draft of the metadata schema policy content.

March - July 2021

Based on the findings from Phase 1, and the specifications for the DiSSCo Policy Self-Assessment Tool defined in French et al., (2021), NHM London collaboratively produced a first draft of a conceptual data model of the metadata schema (see Appendix 1 for the initial version). It is essentially a blueprint for structuring data in the back-end database and interface of the DiSSCo Policy Self-Assessment tool; so that users have a harmonised understanding of policy alignment needs for each service. Moreover, the conceptual data model is an established set of categories with properties and rules that

- 1) define how required policies should be presented,
- 2) define the categorisation of DiSSCo services,
- 3) structure the description of institutional policy documentation, and
- 4) map the institutions' actual policies against service components.

The conceptual data model incorporates the hierarchical system established in the ICEDIG 7.1 policy survey, grouping and describing policies. Policy Category is the highest level of categorisation, followed by Policy Area, and, finally, Policy Component describing a specific element within a Policy Area.

Having the conceptual data model in place helped T2.1 to refine the policy content required for using ELViS. CETAF created a spreadsheet with a tab/sheet for each of the major categories of the data model. The task partners worked systematically on refining the policy categories, policy areas and components provided in the ICEDIG survey (i.e. removing unnecessary policies, adding new policies, regrouping, improving vocabulary, and constraining them with descriptions/definitions). The existing ELViS documentation did not clearly state all the policies needed, and it was sometimes ambiguous as to whether particular policies directly impacted a service, or to what extent alignment was necessary. To overcome these issues, the ELViS coordinators were invited to review the first draft of the metadata schema and provide clarification on certain aspects of policy. The positive feedback and additional information allowed the transition to the next phase of work.

Phase 3: Workshop MS24 'The alignment of institution's policies towards DiSSCo service requirements

September 2021

Milestone 24, '*Workshop to integrate policies and produce harmonisation criteria'* was re-scoped to focus on introducing Task 2.1 work (i.e. the policy metadata schema) to a larger audience within the DiSSCo project and the CETAF working groups in order to:

- Review the policy metadata schema and receive feedback for possible improvements.
- Explore how the metadata schema could be endorsed by the community, to ensure its sustainability beyond SYNTHESYS+.
- Identify challenges of institutional policy alignment towards ELViS, and formulate initial ideas for recommendations on how those challenges could be overcome.

The workshop took place in September 2021 over 2 half days (21st and 22nd). The participants included all SYNTHESYS+ T2.1 partners, CETAF working groups (Collections, Digitisation, Legislation and Publishing), and DiSSCo Prepare Task 7.3 partners. The first draft of the metadata schema was presented and the agenda included a working session to discuss the policy content, including the clarity of the terms used, whether anything was missing, if the level of granularity was sufficient, and any other issues with the schema. The detailed outcomes of the workshop are provided in Tilley, (2021). A summary of the major outcomes regarding improvements to the subsequent version of the metadata schema are:

- Recommendation to improve the names of ELViS service components and definitions to provide more clarity on their links to certain policy requirements.
- No changes were suggested for policy categories and areas.
- Recommendation to add destructive sampling to the Policy Components.
- Recommendation to generalise the Policy Components, clarifying their connection with the institutions to ensure that the policies do not interfere with internal institutional operations.
- Recommendation to clarify essential requirements versus the 'nice-to-haves'.
- Some components appeared to be more technical protocols than actual policies, in particular ELVIS requirements for specimen and data quality.
- Some terminology needed to be improved due to ambiguity.
- In the category 'Institution Policy Area' it was suggested to remove some of the properties and/or not make them mandatory to reduce bureaucracy.

Phase 4 Finalisation of the metadata schema data model.

December 2021

The data model, i.e. the major categories that structure the policy data, were reviewed in light of the work done in phase 2 and phase 3. Only minor changes were made, including the slight renaming of Service Categories as Service Components, renaming 'Service Component Policy Component' as 'Service Policy Mapping'; the removal of service subcomponents and the removal of some properties in the category 'Institution Policy Area' because they were considered not useful in defining the policy requirement for using ELVIS.

Phase 5 Finalisation of policy requirement content

January - April 2022

The finalisation of the metadata version 2 included the modifications from version 1:

- Harmonising the terminology for the ELViS service components,
- The addition of more policy component options to improve clarity of what using ELViS requires.
- Removing policy component's that appeared to be more technical specifications rather than policy

- Asking ELViS developers more specific questions about policies that were not clear from the existing documentation and making alterations based on their feedback.

2.2 Results: Description of the final version of the metadata schema.

The final version of the metadata schema is the foundation for the DiSSCo Policy Self-Assessment tool and its finalised conceptual data model is presented in **Figure 3**. The latter is a graphical representation and overview of the major categories that structure the policy, institutional, and ELViS service information/data (e.g., institutional details and policies; service descriptions and policy requirements), as well as how they relate to each other. Within the DiSSCo Policy Self-Assessment Tool the metadata schema provides a harmonised understanding of which policy requirements are needed for the different aspects of using ELViS. It also provides a structure for curating institution data input via the Policy Self-Assessment tool, making it more interoperable and comparable for analyses, and producing data visuals. The model has been designed to be flexible enough to allow the inclusion of other DiSSCo services beyond ELViS. **Please refer here Metadata schema V2 for definitions of the major categories presented in Figure 3 along with their properties and rules (the full link is provided in the Annex 2).**

The boxes in green across the centre of the diagram serve to classify policies in a hierarchical way, where "Policy Category" is the highest level of categorisation and refers to number of related policy areas; "Policy Area" refers to the classification of a specific policy area or theme. The category "Policy Component" classifies granular policy elements within a policy area. "Policy Component Option" classifies controlled vocabulary terms for options relating to a Policy Component.
The category named "Institution" is for the institution's credentials e.g., name. the category "Institution Policy Area" structures the institution's input information about the current implementation status of relevant policy areas (e.g., whether they have formal policy documentation, and whether the document is public and shareable). The category "Institution Policy Component" classifies the institutions' status of implementation for a specific policy component, which they indicate by choosing from the list of acceptable policy component options related to them. Categories "Service" and "Service Components" classify the details about the DiSSCo service.

“Service Policy Mapping” links specific service components with specific policy components and related policy component options, in other words the level of needed policy alignment.

The following subsections focus on the presentation of the final version content of the metadata schema policy categories (e.g. Policy Category, Policy Area and Policy Component), and Service Component. In general, the metadata schema includes 10 ELViS service components **Table *1***, 2 policy categories **Table *2***, 9 policy areas **Table *3*** and altogether 40 policy components. Due to the large number of policy components, they are not discussed individually, but rather more generally under the different policy areas. For details on the policy components please refer to **Metadata schema V2.**

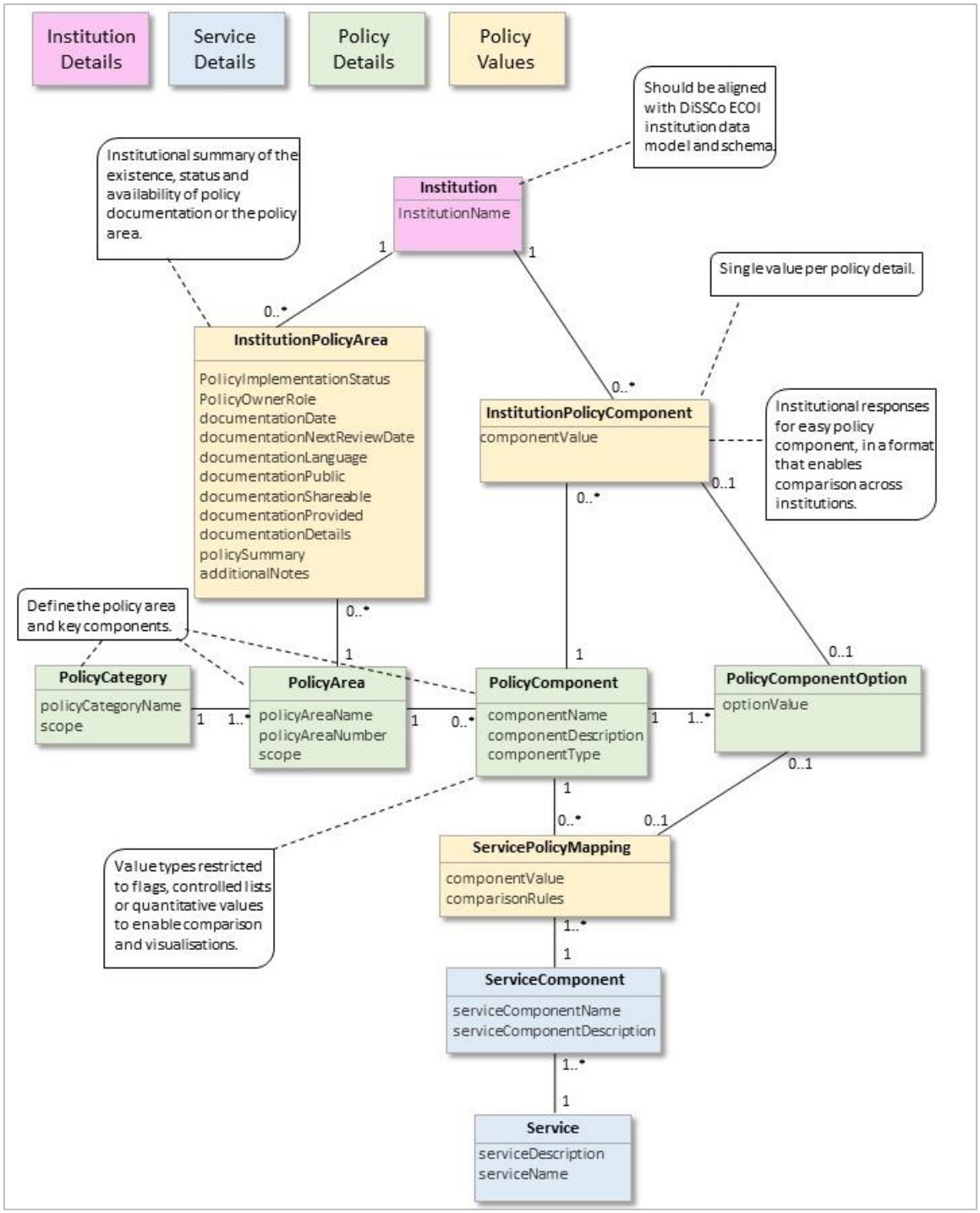


Figure 3. Finalised conceptual data model diagram The data category types (Institutional details are indicated by different colours. The grey writing within the categories are properties that define the type of data input. The solid black lines represent a relationship with other categories. Note: Notations such as 0..1 and 1..* on the relationships indicate the cardinality (nature) of those relationships.

1 = exactly one record: e.g. institutional policy area can only be linked to one institution.

0..1 = zero or one record: e.g. a Policy Component Option does not need to be attached to a Service Policy Mapping but if it does, it can only be attached once.
1..* = one or more records: e.g. Policy Category may be used by one or more Policy Area, and must be used by at least one.
0..* = zero or more records e.g. a Policy Area does not need to be attached to Service Policy Mapping, but if it does, it can be attached to more than one.

2.2.1 Service Component

The service components are defined as specific parts of a service workflow. ELViS service components are presented in **Table *1*** and they relate to specific actions that institutions will undertake to be involved in the service as an access facility/Natural Science Collection host. They were defined based on information provided in the SYNTHESYS+ data managements plan WP6 (Islam and Addink, 2019b). The purpose of breaking a service into components is to facilitate the institution's understanding of why certain policies need to be aligned. It also helps to break down a heavy administrative task of evaluating policy alignment needs that may otherwise seem obscure and overwhelming for some. In addition, not all institutions may want to use every part of a service. For example, not all institutions will be able to provide Virtual Access through Digitisation on Demand.

No policy requirements for using ELViS have yet been identified for the Service Component 'Integrating local collection management systems', but it was included because it will be a future work-flow of ELViS as highlighted in (Islam and Addink, 2019a) and some related policies will potentially be needed in future.

Table 1. ELViS service component

Service Component Name	Service Component Description
Providing people data	This refers to providing people data to ELViS. ELViS will provide detailed information on requesters, TAF admins, and experts associated with institutions' collections (Curators, Collection Managers). People data include: Name, Email, Affiliation, ORCIDs, Researcher and Expertise profiles. People will need to be uniquely identified.

Providing specimen data	This refers to the workflow of providing data on your institutions' specimen holdings via authoritative data sources. ELViS is a transactional platform and so will not store data but rather aggregate it from different authoritative data sources. ELViS will provide information on specimens belonging to institutions, so that users can discover and request them for loans, digtisation on demand and visits. The types of data that will be displayed include: unique persistent identifiers, textual and image data, digitisation level, specimen usage policy as (URL).
Providing collection data	This refers to the workflow of providing data via authoritative data sources on your institutions' collection holdings. ELViS will provide a catalogue of institutional collections, so that users can discover the collections and request them for loan, digitisation on demand or visits. The data that ELViS will display about collections include: unique persistent identifiers, collection holder names, image, collection strength (free text), total specimen counts, digitisation level, Collection usage policy as (URL).
Providing institution/facilities data	This refers to the workflow of providing data via authoritative data sources on your institution and facilities. ELViS requires institutions to provide the following information: Institution names, identifiers, facilities, digitisation priority, policy and workflow.
User Authentication and Authorisation Infrastructure (AAI)	The AAI layer provides authentication (ascertaining that somebody really is who they claim to be / who you are) and authorisation (refers to rules and policies that determine who is allowed to do what / what you are allowed to do)'. Authorisation will be via institutional credentials (or ORCID if institution cannot provide the attributes and be federated) to a DiSSCo service (i.e., ELViS).
Providing virtual access through Digitisation on Demand (DOD)	This service category workflow refers to the requirements needed to fulfil digitisation on demand requests: from dealing with the requesters proposal, the ability to carry-out the digitisation of specimens that are requested, data curation and data publication in a open access repository for aggregation by the DiSSCo infrastructure. A digitisation on demand transaction will include the processing of requestors information (people names, emails, etc.) and transparency on whether the host institution has the capacity, and policies in place to carry out the digitisation. As part of ELViS, institutions will need to provide information on digitisation status, collection citation and research output tracking. ELViS will display this information via an online collections dashboard.

Providing physical access	This service category refers to the policies needed for the for physical visits to collections: receiving the application from the researcher, processing the requesters information including (name, email,), selection of visitors, to be able to provide the physical access requested to an acceptable level considered by SYNTHESYS+ and DiSSCo. As part of ELViS institutions will need to provide information for collection usage monitoring (e.g. numbers of physical access) digitisation status, collection citation and research output tracking. ELViS will display this information via an online collections dashboard.
Providing loans	This refers to the workflow of dealing with loan requests, which involves processing requester information (i.e. names and contact details) As part of ELViS institutions will need provide information for collection usage monitoring (e.g. numbers of Loans) digitisation status, collection citation and research output tracking. ELViS will display this information via an online collections dashboard.
Integrating local collection management systems	This refers to technical requirements for linking ELViS to the institutional collection management systems. API Criteria etc.
Providing destructive sampling	This refers to the workflow of dealing with sampling requests, which involves processing requester information (i.e. names and contact details). These requests may form part of a visit or a loan, or may be independent of these. Any requests will require permissions to be in place in accordance with the Nagoya Protocol, provider country Access and Benefit Sharing Legislation and contractual agreements relevant to the material being sampled.

2.2.2 Policy Categories

The policy details relevant to ELViS have been grouped into two Policy Categories: 1) Data strategy and management; and 2) Physical collections strategy and management. The Policy Category scope and the policy area's that they include are provided in **Table *2***. Unlike the original policy categories outlined in the ICEDIG policy survey, the category for IT strategy is not yet included because the policy alignment needs related to IT are not yet sufficiently known for a category to be defined with reasonable confidence.

Table 2. Policy categories and related policy areas.

Policy Category Name	Scope	Related Policy Area
Data strategy and management	Covers policy areas on data related to people, natural science collections and research, and its governance and stewardship: data accessibility, usability, reuse, shareability, protection from misuse, privacy, standards, quality, publication and attribution.	Data & Digital Media sharing and publication
		Data standards
		Personal data
		Data curation & management
Physical collections strategy and management	Covers policy areas related to care, conservation, compliance, and digital transformation of Natural Science Collections.	Collections loans management
		Collections access & benefit sharing (ABS)
		Responsible Research & Innovation (RRI)
		Physical Collection Management
		Digitisation strategy & prioritisation

2.2.3 Policy Areas and Components

The nine policy areas and their scope are presented in **Table *3***. This subsection provides an explanation of why the policy areas and their policy components and related policy options are needed in relation to the ELVIS service components. **Table *4*** to **Table *12*** present the mapping of service components to the policy components and the rules that indicate at what level they are needed. Some of the needed policy options of the ELViS service components are the same.

Table 3. Policy areas

Policy Area Name	Policy Area Number	Description of Scope
Data & Digital Media sharing and publication	1	Relates to policies covering the releasing of data in published form for external use by others. The area covers FAIR and open data principles, licences, publication of data and digital media linked to Natural Science Collections and research data derived from collections.

Data standards	2	Policies that refer to technical specifications that describe how data should be organised, stored, exchanged, referenced for the consistent collection and interoperability of data across DiSSCo services, systems, users.
Personal data	3	Policies that refer to any information that relates to an identified or identifiable living individual.
Data management and curation	4	This covers the maintenance, security and risk management of collection and specimen data. Data curation is concerned with curating digital specimen and collection data
Collections loans management	5	Refers to policy components related to collection loan management
Collections access & benefit sharing (ABS)	6	Refers to policy components related to compliance with national laws on access and on access to genetic resources, related traditional knowledge and resultant benefit sharing, including Natural History Collection sectoral best practice.
Responsible Research & Innovation (RRI)	7	Policy components referring to responsible research and innovation that anticipates and assesses potential implications and societal expectations with regard to research and innovation, with the aim to foster the design of inclusive and sustainable research and innovation
Physical collections management	8	Policies components related to collection curation, development, housing, environment, care and preservation
Digitisation strategy & prioritisation	9	Refers to policy components that address institutions' strategies, prioritisations and best practices towards digitising their collections.

Policy Area 1: Data & Digital Media sharing and publication

This policy area was formed because ELViS service components have policy alignment needs related to where and how data and media is published and shared. The policy area contains 11 policy components related to data licensing (specimen data, research data, 2D images, 3D images, video/audio recordings), the publication of data to certain open access repositories, FAIR principles, and adoption of the IIIF image standard (see **Metadata schema V2.** for component terms and description). These policy components are requirements for the following ELViS service components: “Providing specimen data”, “Providing collection data”, “Providing institution/facilities”, “Providing virtual access through Digitisation on Demand (DOD)”, “Providing physical access and Providing Loans”. **Table *4*** presents the level at which institutions need to align with the service components required policy components. The defined policy components are based on information provided by documentation and (Islam and Addink,

2019a; Saarenmaa et al., 2019; Islam et al., 2020; Hardy et al., 2020; Hardisty et al., 2020), personal communications with ELViS coordinators and the DiSSCo CSO. More specifically, policy components required for the above service components cover the need for collection, specimen, institutional related data to be published in an open access repository that can be aggregated and displayed by ELViS. This is necessary because as previously mentioned ELViS will be a transactional platform, and so institutions will not be able to provide the data directly to ELViS. The possible implications of institutions not being able to do this will result in ELViS not being able to display information to users about specimens, collections and institutional facilities that are available for access. Additionally, under the current SYNTHESYS+ virtual access and physical access programs, new data on specimens and collections resulting from digitisation or visits should be released to open access repositories, which then can be fed back to ELViS. ELViS does not make it a mandatory requirement, because it is difficult to enforce, that loan requesters should inform the loaning institutions of any derived results, but this is an extra benefit for DiSSCo and in terms of enrichment of specimen/collection data. Policy components related to the default licencing of specimen, institution facilities, 2D images and 3D images and research data have been included because ELViS requires this type of data to have a licencing of *Creative Commons Attribution (CC-BY) at least or ideally Creative Commons Public Domain Dedication (CCO)*(Saarenmaa et al., 2019; Hardisty, 2019). The reasoning for this is so that third parties can access, mine, exploit, reproduce and dis-seminate this data. A policy component expressing the need for institutions to align with FAIR principles was included, it is an obvious necessity because it is at the core of DiSSCo's core and operation, and therefore ELViS. The component called 'Link images to IIIF compatible servers' was added within this policy area but as it has not been made mandatory, we decided to add this as a perceived future requirement since DiSSCo aims to apply the recommendations to use this international image interoperability standard.

Table 4. ELViS service components mapped to the policy components and their options within the policy area Data digital media sharing and publication. Comparison rules indicate to which policy options are acceptable for the service component.

ServiceComponentName	Policy component	Comparisonrule	PolicyComponentOption
Providing specimen data Providing virtual access through Digitisation on Demand (DOD)	Publication of specimen data to an open access repository.	either/or	Geoscience Collections Access Service (GeoCASe) Global Biodiversity Information Facility (GBIF)

Providing physical access Providing loans			Wikidata
			National Biodiversity Network (NBN)
			Global Genome Biodiversity Network (GGBN)
			GenBank
			Europeana
			Institution open access repository
			Other
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Publication of images to an open access repository.	either/or	EUDAT
			Zenado
			Wikidata
			Institution repository
			Other
Providing specimen data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Default license(s) for specimen data	either/or	Creative Commons Public Domain Dedication (CC0) or
			Creative Commons Attribution (CC BY)
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Default license(s) for research data	either/or	Creative Commons Public Domain Dedication (CC0) or
			Creative Commons Attribution (CC BY)

Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Default license(s) for 2D images	either/or	Creative Commons Public Domain Dedication (CC0) or
			Creative Commons Attribution (CC BY)
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Default license(s) for 3D images	either/or	Creative Commons Public Domain Dedication (CC0) or
			Creative Commons Attribution (CC BY)
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Default licence(s) for videos and audio recordings	either/or	Creative Commons Public Domain Dedication (CC0) or
			Creative Commons Attribution (CC BY)
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	FAIR principles	equal to	Yes
Providing specimen data Providing collection data Providing virtual access through Digitisation on Demand (DOD) Providing physical access	Link images to IIIF compatible servers	either/or	Yes
			No
			Not specified

Providing loans			

Policy Area 2: Data standards

The components identified for this policy area are related to description standard requirements for characterising and categorising specimens (and their level of digitisation), collections, and institutional facilities, and include the use of persistent Identifiers for specimens, collections, and institutions (**Table *5***). As mentioned previously, ELViS will incorporate the digital object architecture of DiSSCo that is based on FAIR principles. It has been outlined in the following documentation (Islam and Addink, 2019b; Islam and Addink, 2019a; Hardisty et al., 2020; Addink and Hardisty, 2020) as well as during personal communication between L. Tilley and ELViS coordinators, that, the digital objects will be supported by community recognised standards at the specimen level, which are ABCD (and its extensions) and DarwinCORe, and at the collection level the TDWG collections Description (CD). Therefore, we provided the requirements for collection and specimen descriptions as two separate policy components.

Table 5. ELViS service components mapped to the policy components and their options within the policy area Data standards. Comparison rules indicate to which policy options are acceptable for the service component.

ServicePolicyComponent Name	Policy component	Comparisonrule	PolicyComponentOption
Providing specimen data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Specimen data description standards	either/or	ABCD-EFG (for geological objects)
			Darwin Core
			ABCD (for biological objects)
			ABCD-DNA
		Equal to	MIDS
Providing collection data	Collection description granularity	Equal to	Yes

Providing specimen data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Collection data description standards	either /or	TDWG Collection Description Data Standard
Providing specimen data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Unique persistant identifier for specimens (non digitised and digitised)	either/or	Cetaf stable identifiers
			Natural Science Identifiers (NSid)
			DOIs
			Other
Providing specimen data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Unique, persistant identifiers for collections.	either/or	GRSciColl (GBIF registry of scientific collections)
			DOIs
			Other
			Non-Specified
			None
Providing Institution/facilities data	Unique persistant Identifiers for institutions	either/or	GRID (Global Research identifier Database)
			ROR (Research Organisation Registry Community)

The Minimum Information about a Digitised Specimen (MIDS) standard [Minimum Information about a Digital Specimen (MIDS) - TDWG]), will be used for characterising the level of digitisation of the Natural Science Objects that ELViS will display. ELViS does not require institutions data to meet a certain level of MIDS, but it does need institutions to characterise the digitisation of their collections and specimens using MIDS. In the schema we included the requirement of MIDS as an option within the policy component ‘Specimen data description standards’ but it should be removed and implemented as its own policy component. The

reasoning for this is because MIDS is a specific necessary policy alignment need, whereas the other options in the list can be either or. Moreover, those different levels of requirements within a policy component make it difficult to visually display percentage of institutions that align with the component as whole, because some may not have MIDS, but comply with the repositories. This difficulty became apparent from displaying the results of the mapping exercise and reported from DPP T7.3 in the development of the policy tool (see section 4 for recommendations).

The rationale behind the inclusion of the policy component ‘Collection description granularity’ is derived from the fact that, as mentioned in Policy Area 1, ELViS will aggregate collection data from the CETAF registry of collections, which follows the collection classification scheme developed under SYNTHESYS+ (Tilley et al., 2019). This classification scheme follows the TDWG CD and will underpin the Collection Digitisation Dashboard that will be implemented into ELViS as a first window into collection discovery. For the users to discover collections for access in the hosting institutions, it is important that institutions should provide at least the lowest level of granularity (most general) of data about their collections in accordance with the Collection Digitisation Classification scheme.

Unique persistent identifiers (PIDS) are needed to resolve digital specimens/objects, institutions, physical collections and specimens, unambiguously online via DiSSCo services (Hardisty, 2019; Islam and Addink, 2019b; Hardisty et al., 2020; Hardisty et al., 2021). From the existing documentation and discussions in other SYNTHESYS+ work packages, we could identify with confidence that ‘ROR (Research Organisation Registry community)’ identifiers will be required for the identification of institutions. However, it is unclear what types of Persistent identifiers will be required for specimens and collections, so in the metadata schema we listed the ones commonly mentioned and that have wide community acceptance. During the later stage of the task and after CETAF attended a physical DiSSCo CSO meeting (Leiden, July 31st, 2022) it became apparent that PIDS requirements for Digital objects should be removed. Institutions will need to provide only unique identifiers for the physical specimens. This is due to new information that DiSSCo will allocate a Natural science identifier (NSID) for a digital specimen. A digital specimen is a digital representation of a physical specimen and contains data about it, the digital specimen can be extended to include other data derived from the physical specimen such as linkage to research papers, data sets, different images, chemical data, genetics etc. The NSID is provided to the Digital Specimen and all that it contains. A physical specimen PID

such as the CETAF stable identifier is required at the institution level to keep a linkage between the digital and physical specimen (Hardisty, 2019; Islam and Addink, 2019b; Hardisty et al., 2020; Hardisty et al., 2021).

Policy Area 3: Personal data

Since being a part of ELViS will involve institutions to handle personal data of users and own staff, the policy area Personal data was defined. Personal Data contains two policy components (**Table *6***), one refers to the General Data Protection Regulation (EU) 2016/679 (GDPR) and the other to the use of 'identifiers for person Authentication'. It is indicated with confidence that GDPR is a required policy for most of the service components because it is an EU level requirement that must be followed when capturing and processing people data in Europe, and these service components involve the use of personal data. The policy component 'identifiers for person authentication services' is needed for the service components Providing personal data and the User Authentication and Authorisation Infrastructure (AAI), and will be used to control access to certain services and data (Addink et al., 2020; Islam, 2022), moreover, to manage the different levels of access by linking users to roles and their entities. The acceptable options for people identification in ELViS is either institution credentials which will be linked to the service via eduGAIN, or ORCiD identifiers. Investigations are still on going in the use of institutional credential via eduGAIN (Islam, 2022). Since the AAI is still being developed it is likely that there will be future policy requirements around privacy and authentication management.

Table 6. ELViS service components mapped to the policy components and their options within the policy area Personal data. Comparison rules indicate to which policy options are acceptable for the service component.

Service Policy Component Name	Policy component	Comparison rule	Policy Component Option
Providing people data User Authentication and Authorisation Infrastructure (AAI)	Identifiers for person authentification	either/or	ORCID IDs
			Institution credentials
Providing specimen data User Authentication and Authorisation	General Data Protection Regulation (EU)	Equal to	Yes

Infrastructure (AAI) Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	2016/679 (GDPR)		

Policy Area 4: Data management and curation

The policy components under this policy area are 'Data management plan', 'Protection of sensitive data' and 'Information incident management processes defined' (**Table 7**). These policy components come from more general needs of DiSSCo as mentioned in Saarenmaa et al., (2019); Hardisty et al., (2020) and feedback for ELViS coordinators, and have been identified as requirements for all service components since they all deal with data in some aspect and needs different levels of protection.

A data management plan (DMP), as mentioned in Saarenmaa et al., (2019) and references therein outlines the questions and approaches surrounding data management and, when applicable, within the scope of a funded research project. A DMP could be either a project based or an institutional one. Specifically, regarding the policy component 'Data management plan', ELViS coordinators consider that it is important for the function of ELViS that institutions should have their own data management plan in place because this will be the basis for DiSSCo to create Service Level agreements. The policy component 'Protection of sensitive data' refers to sensitive data not covered by GDPR. ELViS' requirement for institutions to have a policy that covers this component comes from DiSSCo's mandate for data to be as open as possible and as closed as necessary, with specific legislation applying to DiSSCo data based on international conventions related to natural sciences (e.g. CBD, CITES, Nagoya protocol on access and benefit sharing), with supranational legislation (e.g. European Union's INSPIRE Directive, Habitats Directive) and with contracts bearing on the data (Saarenmaa et al., 2019; Hardisty, 2019; Hardisty et al., 2020). The policy component 'Information incident management processes defined' was also included based on the above mandate, in that institutions should have policies to deal with data breaches that may break the above legislations.

Table 7. ELViS service components mapped to the policy components and their options within the policy area Data management and curation . Comparison rules indicate to which policy options are acceptable for the service component.

Service Policy Component Name	Policy component	Comparison rule	Policy Component Option
Providing specimen data Providing collection data Providing institution/facilities data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Data management plan	Equal to	Yes
Providing specimen data Providing collection data Providing institution/facilities data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Protection of sensitive data	Equal to	Yes
Providing specimen data Providing collection data Providing institution/facilities data Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Information incident management processes defined	Equal to	Yes

Policy Area 5: Collections loans management

Only one policy component (**Table *8***) has been identified under this policy area so far, and this is "Incoming and outgoing loans". Institutions will generally have loan policies or processes dictating the actual loan process. This component is only relevant for the service component "Providing loans". As far as ELVIS is concerned, the only real specified requirement mentioned for institutions regarding loans is that they should record incoming and outgoing loans, this is so

that the usage of institution collections can be reported via the Collections Digitisation Dashboard that will be embedded in the ELViS platform (Islam and Addink, 2019a). But obviously, every institution using ELViS for processing all or parts of the loan process needs to implement that in their institutional loan policy. Some of the ELViS user stories provide indications that institutions should have a loans policy. For instance, collection managers need to see outstanding and current loan status within ELViS, and to review past loan requests, and for researcher/requesters to ensure they can comply with loan terms before submitting a loan request.

Table 8. ELViS service components mapped to the policy components and their options within the policy area Collections loans management. Comparison rules indicate to which policy options are acceptable for the service component.

Service Policy Component Name	Policy Component	Comparison Rule	Policy Component Option
Providing loans	Incoming and outgoing loans	Equal to	Yes

Policy Area 6: Collections access & benefit sharing (ABS)

The components under the Collections access & benefit sharing (ABS) are needed for 5 ELViS service components (**Table *9***): “Providing specimen data”, “Providing virtual access through Digitisation on Demand (DOD)”, “Providing physical access”, “Providing loans” and “Providing destructive sampling”. The components refer to the requirements outlined under the relevant supranational (e.g., EU) and national laws pertaining to Access and Benefit Sharing including the Nagoya Protocol on Access to Genetic Resources and Benefit Sharing. It is important that institutions have some policy in place that reflects national or supranational (e.g., EU) regulations on access to genetic resources for their utilisation, sharing of the benefits arising from such utilisation, especially through Commercial use, and compliance with national ABS-legislation of countries that provided these genetic resources. Not having policies on ABS in place would lead to user’s/researchers who access the collections, maybe not being to publish their data acquired, as well as facing sanctions depending on the laws in the country of the host institution and of the researcher. This could have a knock-on impact on ELViS and DiSSCo as a whole in terms of not being able to use the data and increases risks relevant to legal compliance and ethical and moral reputation.

Table 9. ELViS service components mapped to the policy components and their options within the policy area Collections access and benefit sharing (ABS). Comparison rules indicate to which policy options are acceptable for the service component.

Service Policy Component Name	Policy Component	Comparison Rule	Policy Component Option
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans Providing destructive sampling	Utilisation of Genetic Resources	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans Providing destructive sampling	Benefit sharing	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans Proving destructive sampling	Traditional Knowledge	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans Providing destructive sampling	Commercial use	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans Providing destructive sampling	Access to information and associated data on genetic resources	Equal to	Yes

Policy Area 7: Responsible Research & Innovation

This policy area was defined based on the policy requirements of the European Commission for institutions involved in the EU research funding programs. The term Responsible Research & Innovation (RRI) is the term coined for in the Horizon 2020 policies “to encourage societal actors to work together during the whole R&I process to better align R&I with the values, needs and expectations of society”. Responsible research and innovation is promoted via: public engagement, open access, gender, ethics, science education, and integrated actions that for example promote institutional change European Commission, Directorate-General for Research and Innovation, 2017; Pain, (2017). The components presented under this area (**Table *10***) are based on the requirements currently needed by SYNTHESYS+ to be a part of the TA and VA. The EU commission requires a gender plan, ethics and public engagement in order to require EU funding, thus T2.1 considered this to be important for future funded calls for Transnational access and Virtual access.

Table 10. ELViS service components mapped to the policy components and their options within the policy area Responsible research and innovation. Comparison rules indicate to which policy options are acceptable for the service component.

ServicePolicyComponentName	PolicyComponent	ComparisonRule	PolicyComponentOption
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Public Engagement	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Gender equality plan/strategy	Equal to	Yes
Providing virtual access through Digitisation on Demand (DOD) Providing physical access Providing loans	Ethics	Equal to	Yes

Policy Area 8: Physical collections management

Four policy components (**Table *11***) have been identified under Physical collection management: “Visitor access”, “Recording visitor numbers”, “Culturally sensitive objects (e.g., from

anthropology)" and "Destructive sampling". They are required for the service components "Providing specimen data", "Providing physical access", "Providing loans", and "Providing destructive sampling". It was unclear what policies on physical collection management needed to be aligned towards ELViS at this point. But the reasoning for adding those so far are as follows.

The policy component "Visitor access" refers to institutions having a policy that governs visitor access. This is considered a relevant policy based on the current requirements of institutions that wish to take part in the SYNTHESYS+ Transnational Access calls, in which they have to do a CSAT assessment (Collections Self-Assessment Tool (CSAT) - ppt download (slideplayer.com). It is also considered because ELViS is supposed to help in the arrangement of visits to collections, so using ELViS should be reflected in the institution's visitor access policy. T2.1 partners reviewed the CSAT requirements; it appeared that some of them were too detailed for what was possibly needed for using ELViS as it covered a lot of areas beyond collection access *per se*. T2.1 partners decided to add a general policy component requesting that institutions using ELViS have a policy on receiving visitors and ELViS itself should direct potential users to the host's own visiting terms. This would assist ELViS as a gateway of providing access to collections and ensure that institutions could advertise their willingness to receive visitors. The policy component "Culturally sensitive objects" was included not only for the same reasons as including the "Visitor access" policy component, i.e., to explain the context of access to these objects, but also to make sure that every sensitive object is handled with appropriate care.

Recording visitor numbers is a clear need for institutions and ELVIS for reporting both use of institutional collections (Islam and Addink, 2019a). The policy component "Destructive sampling" was included as a leading element of a new service component "Providing destructive sampling". This became apparent in the discussion of the MS24 milestone workshop (Tilley, 2021), as well as being mentioned in the ELViS user story requests, for example, curators handling of destructive handling requests.

Table 11. ELViS service components mapped to the policy components and their options within the policy area Physical collections management. Comparison rules indicate to which policy options are acceptable for the service component.

ServicePolicyComponent Name	Policy component	Comparisonrule	PolicyComponentOption
Providing physical access	Visitor access	Equal to	Yes

Providing physical access	Recording visitor numbers	Equal to	Yes
Providing physical access Providing loans Providing destructive sampling	Culturally sensitive objects	Equal to	Yes
Providing physical access Providing destructive sampling	Destructive sampling	Equal to	Yes

Policy Area 9: Digitisation strategy & prioritisation

Policy components under this policy area (Digitisation strategy, minting of digital specimens with unique persistent identifiers, prioritisation criteria) in relate to ELViS needing to know the digitisation capabilities of institutions, as not all institutions will be able to provide digitisation on demand, and those that can will be able to provide at different capacity levels (Islam and Addink, 2019a). There needs to be transparency in who can provide what and how, and what are the institution's priorities for digitisation. From the design document of ELViS (Islam and Addink, 2019a) it appears that there will be a tier levelled service for institutions. A tiered service provision level will be needed in which some institutes can provide a higher provision level than others. To encompass these needs the policy components in **Table *12*** were added. The policy options for the policy component “Prioritisation criteria” was based on the options provided in the ICEDIG 7.1 survey and Bakker et al., (2018). In light of new insight provided by the CSO (Laura Tilley personal communication Leiden July 2022), it is recommended that ‘Minting of Digital Specimens with Unique Persistent identifiers will need to be removed because it will be DiSSCo that mints the Digital objects rather than the institution, as mentioned above the physical collection/specimen needs a resolvable PID that allows the physical object to be linked to the extended digital object.

Table 12. ELViS service components mapped to the policy components and their options within the policy area Digitisation strategy & prioritisation. Comparison rules indicate to which policy options are acceptable for the service component.

Service PolicyComponent Name	Policy Component	Comparison Rule	Policy Component Option
Providing Institution/Facilities data	Digitisation strategy	Either/or	Oversight and governance
			Digitisation goals and

			prioritisation
			Digitisation workflows
			Digitisation capacity
			Mass digitisation
			2D imaging
			3D imaging
			DNA sequencing
			Digitisation data standards
			Other
	Minting of Digital Specimens with Unique Persistent identifiers	Equal to	Yes
	Prioritisation criteria	Either/or	Institutional science strategy
			National science strategy
			Collections development
			Collections moves
			Collections conservation
			Digitisation workflow effectiveness
			Public engagement
			Institutional research focus
			Global research focus
			External funding
			Compliance with standards
			Object identifiers
			Cultural importance
			Relevance for economic activities
			Relevant for fundamental research
			Relevant for access to primary biodiversity data
			Research focused on studying

			processes and trends (e.g. evolution)
			Contribute to conservation (policy)
			National research focus
			Institutional revenue
			Education purpose
			Financial donations
			Other

3. Case study of mapping institutional policy data

The case study entailed mapping the T2.1 partners' policy information to the metadata schema, and the two reasons for doing so were: 1) To test the practicality of the schema in terms how easy it is to understand vocabulary and description of the policy areas and components, and to test institution policy documentation categories specifically designed for the mapping. 2) To provide a preliminary analysis of the current state of natural science institutions policy alignment towards the currently defined ELViS service requirements (i.e., where are the policy gaps, which policy areas are most aligned). 3) To assess whether the user interface questions linked to the policy components are understandable and reflective of the intent of the policy components. This analysis was important for identifying potential challenges for institutions to align their policies with DiSSCo service requirements, and to provide future recommendations to overcome these possible challenges.

Two iterations of the case study were done in September 2021 and in May 2022. The first was mostly for testing the practicality of the policy component content of the metadata schema V.1, but also an attempt to identify challenges such as institutions policies that prevent alignment, policy gaps across all institutions, etc. This was done in time to be discussed in the MS24 workshop (Tilley, 2021). The second iteration of the case study was conducted using the finalised version of the metadata schema. The results from the first iteration were reviewed by T2.1 partners and used to improve the metadata schema. The second iteration of the mapping exercise is based on the final metadata schema, in which the results are presented here.

3.1 Methodology

The mapping exercise was conducted in a google sheet (Mapping institution policies to metadata schema V.2 2022) that contained 4 sheets for data input. The first sheet "InstitutionsPolicyComponent" contained all the policy components and questions relating to the policy components. Each institution participating in the case study had their own designated column where they could indicate which components that they have or don't via tick boxes or drop-down answers (yes, no or not specified). The following 3 pages were created for collecting information about participating institution's policy documentation for every policy area, broken down by the properties of the category 'Institution Policy Area'. CETAF mapped all the partners relevant policy data, which was initially collected at the start of the T2.1 via the ICEDIG policy survey, to the final metadata schema. Institutions were then asked to check through their policy information, and update where necessary.

Selected raw results from the exercise were transformed into visual representations (bar charts and pie charts) to easily analyse

- the number of institutions that align with ELViS policy components in the final metadata schema within each policy area,
- the implementation status of each of the policy area (i.e., whether it a formal documented policy, undocumented procedure, partly documented or there is no policy in place), as well as
- for those institutions that do have documentation, how many can share it publicly.

Charts were created to show the state of alignment of policies per ELViS service component. In addition to organising the raw data provided, CETAF also reviewed the policies documentation that was provided to see if there was further qualitative insight that could be gained on the policies, for instance whether the policy was defined at a governmental or institutional level, to what extent the policy documentation covers the policy components. Although, not all policies were written in English which made comprehension challenging and interpretation may not be as accurate as it would be for a native speaker of a given language. CETAF also reminded and encouraged institutions to provide the data for the exercise and helped highlight areas that may

have been incomplete. The results of the analyses presented below. T2.1 partners were asked about the experience of conducting the mapping exercise during a tasking meeting.

3.2 Results

3.2.1 Practicality of using the metadata schema

Not all participants were present at the meeting to provide feedback, but those that attended generally thought that the policy component names and descriptions were clear. However, within the mapping exercise sheet the ELViS service components were not presented in relation to the policy components and sometimes partners found it hard to answer the components without any context as to why they were needed or relevant. With regards to providing information about policy documentation, some institutions said that some of their policies have a mixture of policy areas proposed here within one document, or elements that do not all fall under the policy areas as defined by the metadata schema. This is a difficult thing to harmonise because each institution has a different need of naming and grouping elements for policy areas proposed here. Institutions found the question related to the policy component 'incoming and outgoing loans not representative (Does your institution follow a documented loan policy?).

3.2.2 Level of institutional policy alignment.

The results of this task rely on the completeness of the policy data that the institutions have provided, and this varies for different institutions, some institutions were more responsive than others for unknown reasons. This being said, the completeness is good enough for providing insightful results and interpretation (**see google sheet** 'Mapping institution policies to metadata schema V.2 2022' **for raw data, and questions)**. Regarding the completeness of the three sheets for the Institution Policy Area category, which is for collecting information about participating institution's policy documentation, all institutions provided data for the property "documentation implementation status" (formal policy document present, undocumented procedure, partly documented, no policy in place) for each of the policy areas. But not many institutions provided information about the point of contact for the policies, the date of renewal or date of documentation. Also, even though the institutions may have said they have documentation, not all provided information on whether it was public or shareable or provided a working link or information about the document. The reason for the information is unknown.

Overall, there are no policy areas that have no alignment at all, but the level of institutional alignment with the subordinate policy components varies a lot. The policy areas where most of the policy components show high institutional alignment levels, with over half of institutions aligned, are policy area 1 “Data & digital media sharing and publication”*;* policy area 6 “Collection’s access and Benefit sharing”, policy area 7 “Responsible Research & Innovation (RRI)”, and policy 5 “Collection loans management”. Below, we describe the level of alignment within each policy area in more detail. See Appendix 3 for graphs should alignment of policy components per service component.

Policy Area 1: Data & Digital Media sharing and publication

The component mapping exercise shows that alignment of the nine participating institutions is high with the policy components needed for publication of the different types of data (specimen data, images, collection data, and institutional facilities data) to open access repositories (**Figure 4**).

All nine institutions publish their **specimen data** to GBIF. One and four institutions retrospectively also provide their specimen data to the National Biodiversity Network (NBN), and GGBN. Six institutions have their own public repository. For **images** most institutions publish data on their own open access repository, three of them additionally to Zenodo. Three institutions (also) use other repositories not listed in our case study. Seven institutions provide their **facilities data** to the CETAF registry of collections, four institutions (also) use other repositories not listed in our case study. The same seven institutions provide their **collections data** to the CETAF registry of collections, four of them additionally to the GBIF registry of scientific collections, two institutions (also) use other repositories not listed in our case study, one uses Wikidata and another an institutional repository.

There is variability in the alignment of default licensing for required data types (research data, specimen, 2D and 3D images, and video and audio recordings) The alignment is high for specimen data with 7 institutions having at least the minimum. Four institutions such as NHM Vienna have an array of different default licences for specimen data so not all of the specimen data is open access. The alignment with anticipated requirements for Default licences for research data, 3D images and video’ and audio recording is lower. Over half the institutions follow FAIR principles.

The component with the lowest alignment is ‘Link images to IIIF compatible servers, only 1 institution says that they align.

All but one institution has formal policy documentation related to “Data and digital media sharing and publication”. Only 2 institutions have their documentation publicly available but 4 say that they can share it publicly, two provide no information regarding the accessibility or shareability. On further investigation of the content of the documentation available (from 4 institutions) the documentation does not enclose information related to all the components within the policy area. For instance, CSICs policy documentation mainly focuses on their institution's open repository for research data but it has another document ‘recommendations for providing FAIR data’. NHM Vienna focuses on trademarks and copyrights but does not provide details about types of licences, repositories, or FAIR data. BGM follows the Flanders policy which is considered liberal but does not provide clarity on the level of access to research data, in addition they have guidance in the data management plan for making data openly accessible, but this does not cover photos, archives or books. The Royal Botanic Gardens Kew provides the most comprehensive with regards to licences. None of the policies directly mention the publishing of specimen data collection data to external repositories.

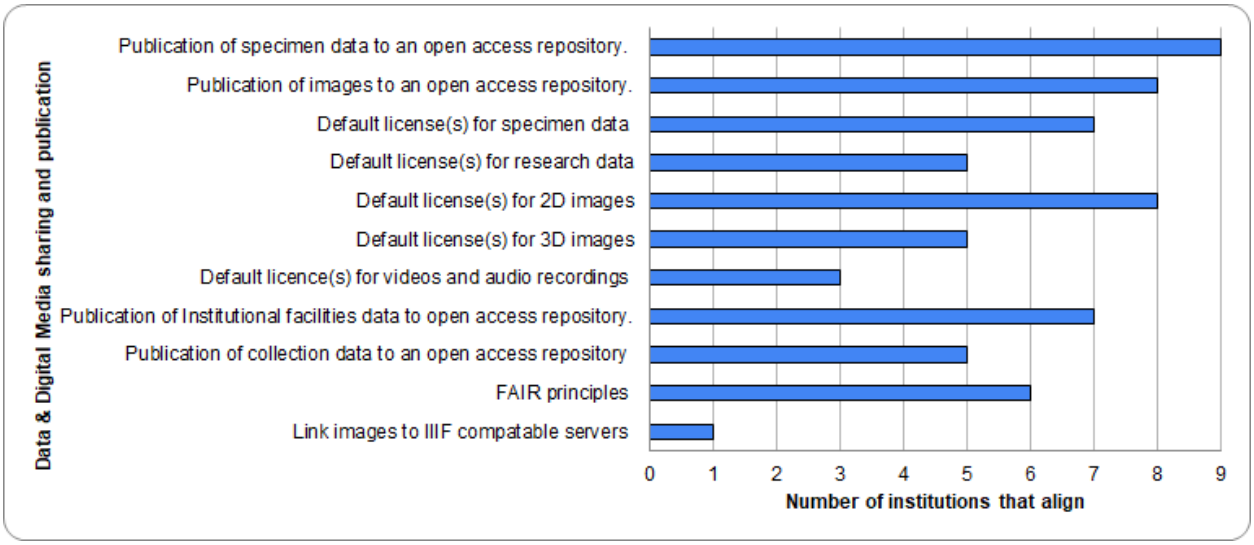


Figure 4. Bar chart showing the number of institutions that align ELViS policy component requirements within the Data & digital media sharing and publications. Total number of institutions = 9.

Policy Area 2: Data standards

Alignment toward policy components is generally low for this policy area (**Figure 5**). The policy component with the most alignment is Specimen data description standards generally, but the alignment in the use of MIDS is low.

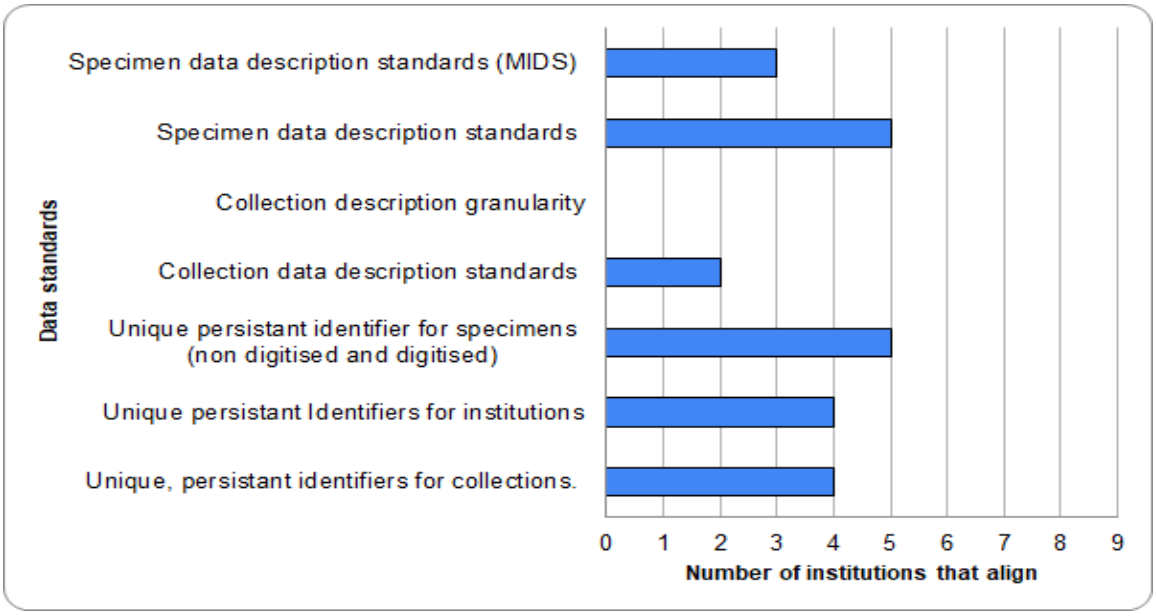


Figure 5. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Data standards. Total number of institutions = 9.

Policy Area 3: Personal data

Alignment with GDPR is 100% (**Figure 6**)**.** There is low alignment in the policy component 'identifiers for person authentication'. The mention of Authentication via institution credentials and/or ORCIDs is not mentioned in the few institution documents that could be observed.

Seven institutions have a document on personal data (**Figure 7**). For most of them this is a national or EU level documentation, rather than one specific documentation for the institution. RBGK appears to have the most comprehensive GDPR policy out of the institutions where their policies can be viewed, it appears to align with the EU GDPR guidelines.

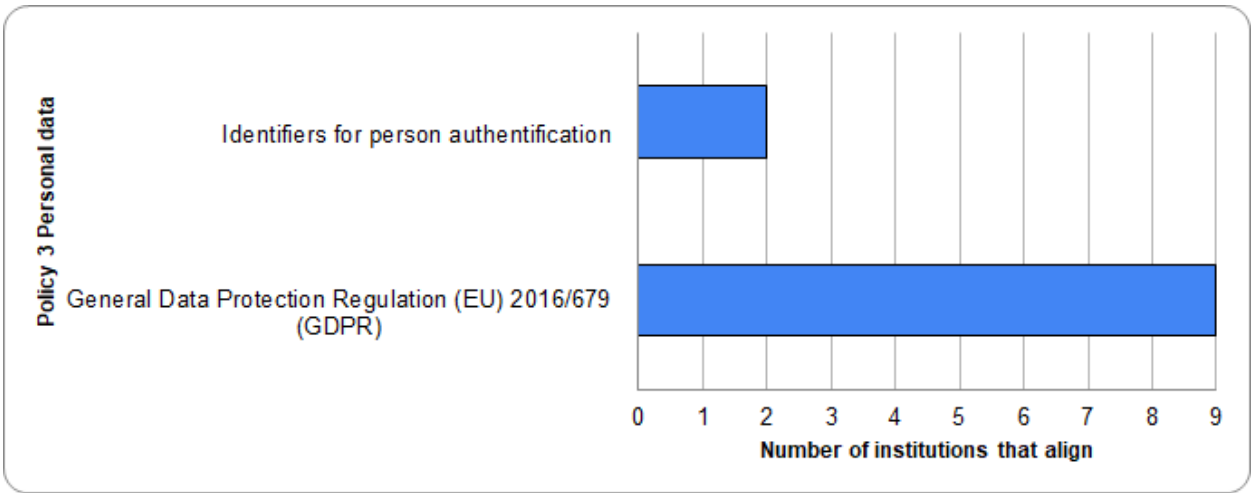


Figure 6. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Personal data. Total number of institutions = 9.

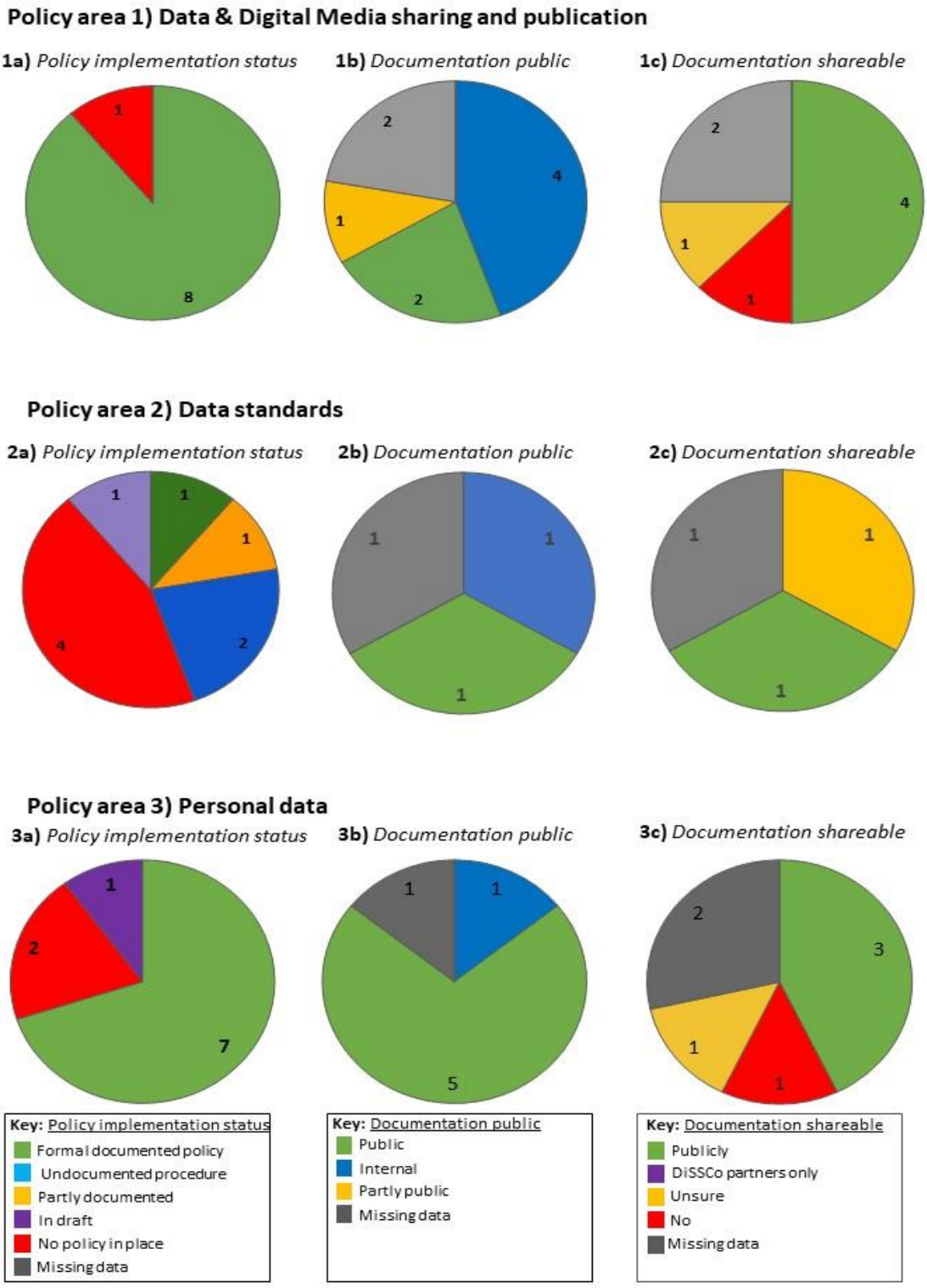


Figure 7. Pie charts presenting the Policy implementation status, Documentation Public and shareability for Policy areas: 1) Data & digital media sharing and publication, 2) Data standards, 3) Personal data. Number inside the pie charts represent the number of institutions.

Policy Area 4: Data management and curation

Half or fewer of the institutions say that they align with the component requirements for the Policy Area, Data management and Curation (**Figure 8**). Four of the institutions said that they had formal policy documentation in place but only one said it was sharable. However, there is some discrepancy in the data provided for this policy area. Within the institution component mapping part of the exercise, seven institutions (**Figure 11**) said that they followed procedures on the protection of sensitive data and three for the information incident management process, although they do not indicate that they have it as an unwritten procedure. The institution that said they can publicly share their documentation did not provide it, so it is difficult to see if it can be provided.

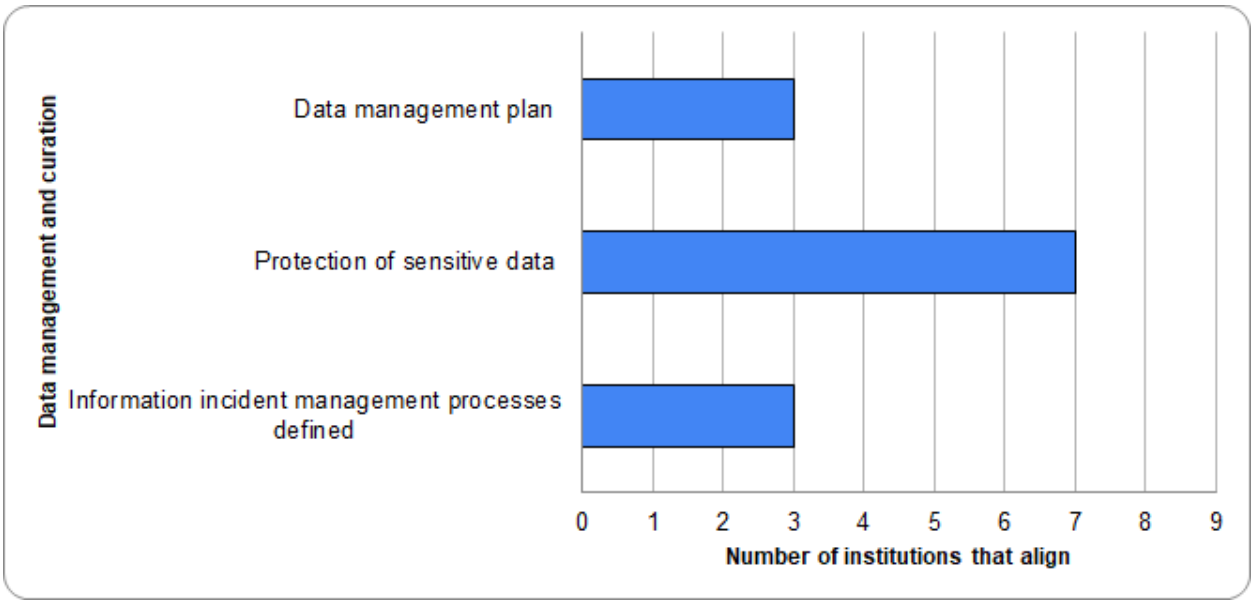


Figure 8. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Data management and curation. Total number of institutions = 9.

Policy Area 5: Collections loans management

Six institutions stated that they record incoming and outgoing loans (**Figure 9**). Eight institutions indicated that they have a formal document policy on collection loans management. Of the institutions who have formal policy documentation, four stated that they could publicly share it with one able to share only partly to just DiSSCo partners (**Figure 9**). RBINS, CSIC, RBGK and NHMW were the only institutions that provided an accessible link to their policy documentation.

Amongst them there is a degree of variability as to whom the policy applies. For instance, CSICS policies are written for the loan requester rather than the internal staff who manage the loan for the requester. RBINS, RBGK and NHMW provides a comprehensive policy on Loans management for providing and acquiring, and more specifically it includes how to record outgoing loans, the timeframe and the number of loans that can be requested by one person. Additionally, RBINS and RBGK state if the loan request results in any publication the institution must be cited, copies of the publication and photos must be provided to the museum, and the loaning institution must be informed of any changes to the museum (including new taxonomic diagnosis).

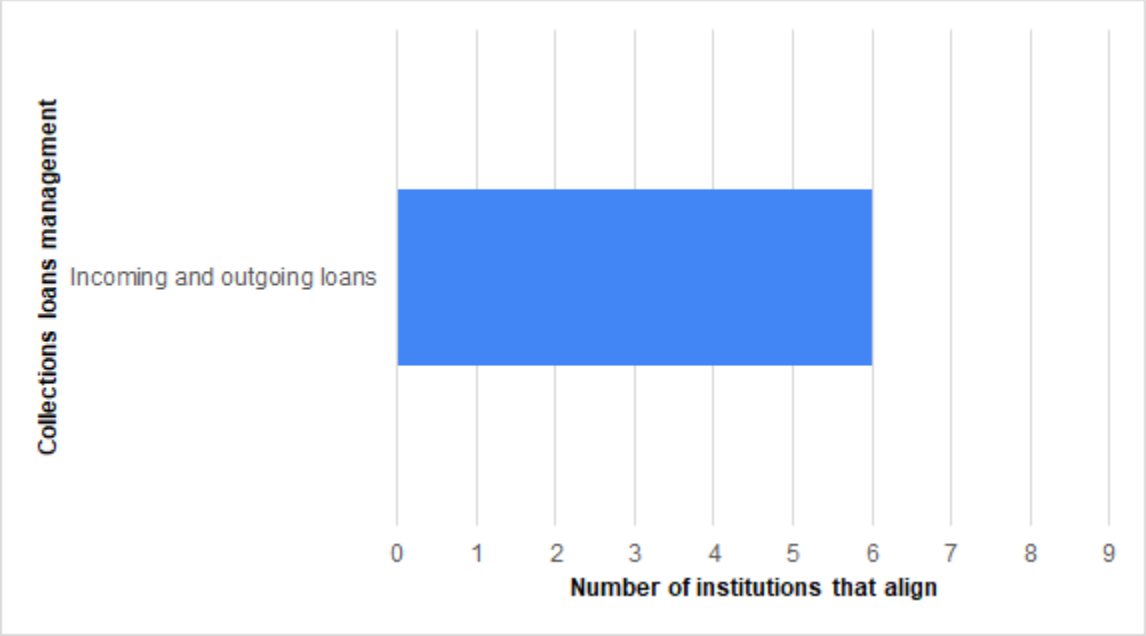


Figure 9. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Collections loans management. Total number of institutions = 9.

Policy Area 6: Collections access & benefit sharing (ABS)

All institutions say they align with regards to the “Acquisition of genetic resources and prior information” and “Access to information and associated data on genetic resources”(**Figure 10**). Seven institutions have a formal documented policy, some follow national level documentation (CSIC, NHM Vienna, MNHN) and the rest have institutional documentation that reflects the national level. There are data gaps due institutions not fully completing the google sheet, and

not all the documents could be accessed. RBGK provides comprehensive documentation at the institutional level.

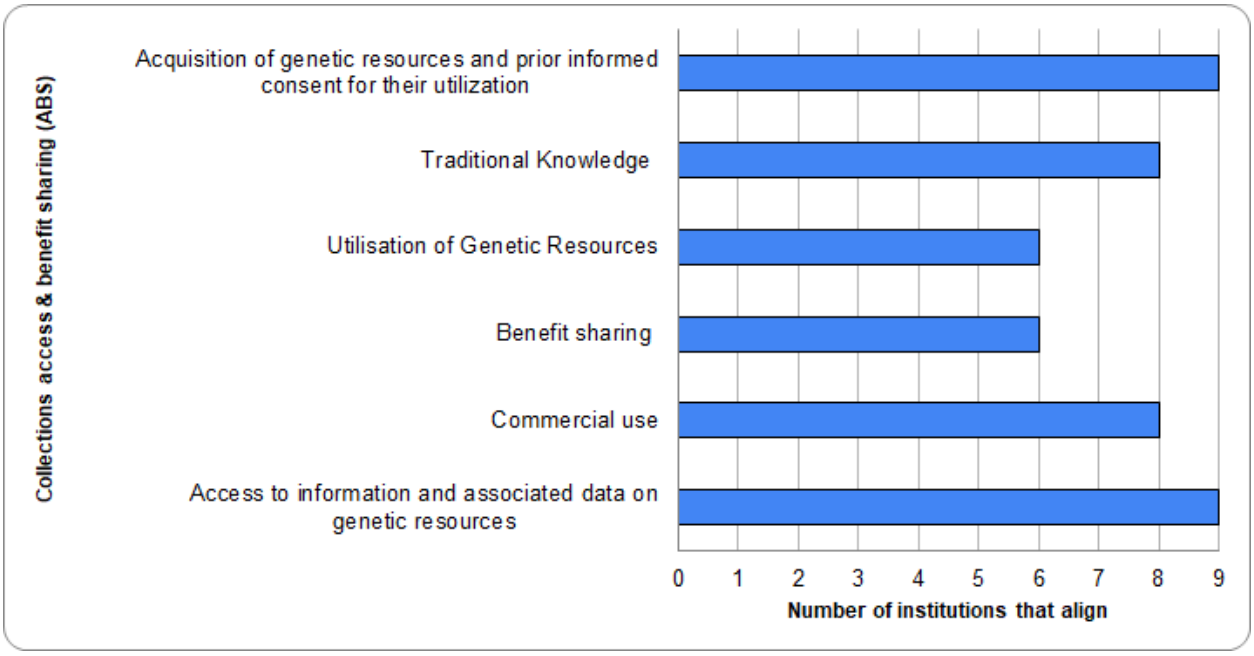


Figure 10. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Collections access & benefit sharing (ABS). Total number of institutions = 9.

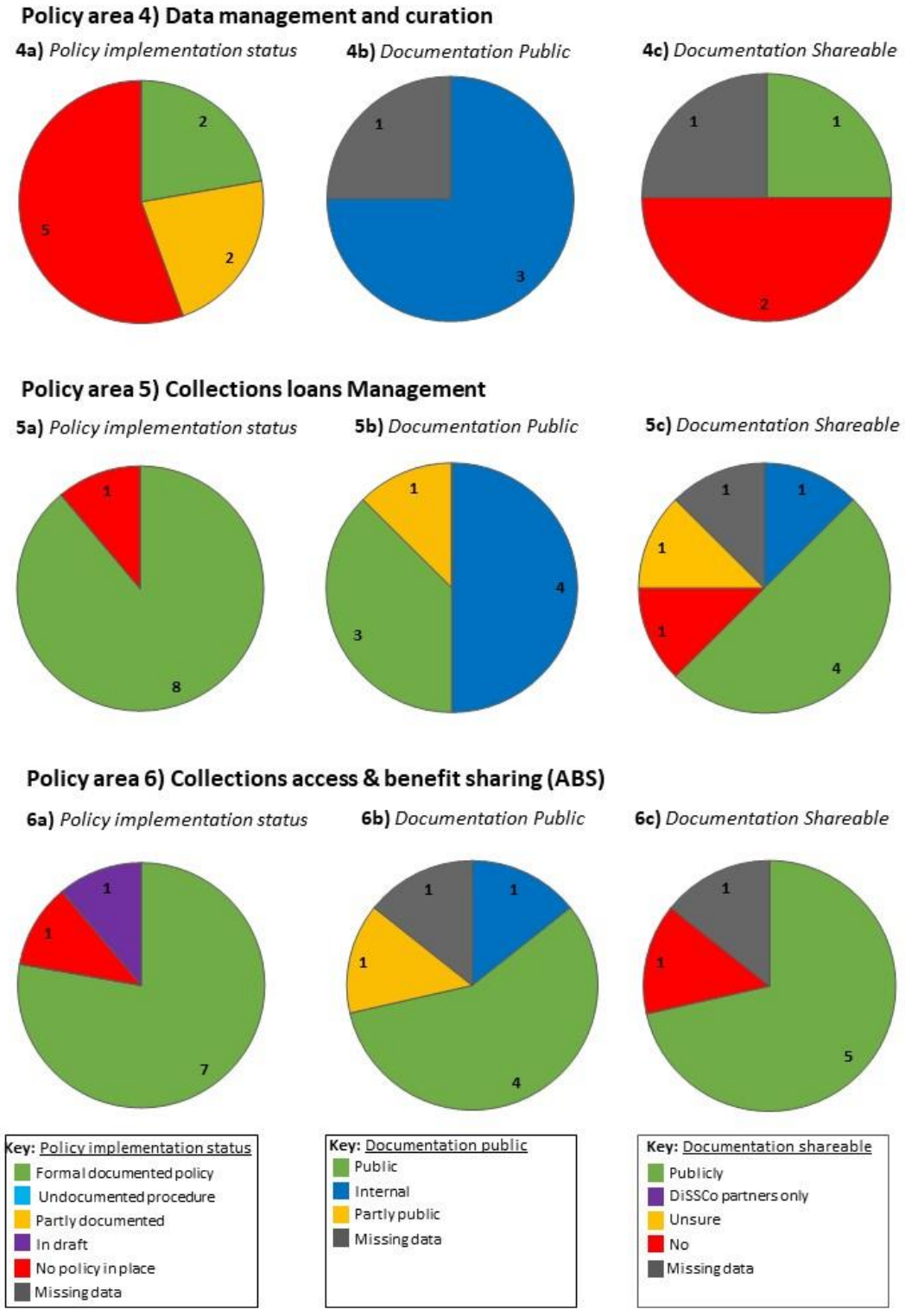


Figure 11. Pie charts presenting the Policy implementation status, Documentation Public and shareability for Policy areas: 4) Data management and curation, 5) Collection loans management ,

6) Collection access & benefit sharing. Number inside the pie charts represent the number of institutions.

Policy Area 7: Responsible Research & Innovation (RRI)

There is high alignment by all or nearly of the institutions according to the survey for a gender equality plan and Ethics (**Figure 12**). Alignment with the policy component on public engagement is lower. The institutional plans appear to mostly follow national legislations, BGM follows CETAF and Horizon 2020 which may be outdated. It is difficult to see the exact compliance with the gender equality plan due to only a small number sharing their documentation. There is a need now to have a gender equality plan that follows certain criteria of the EU.

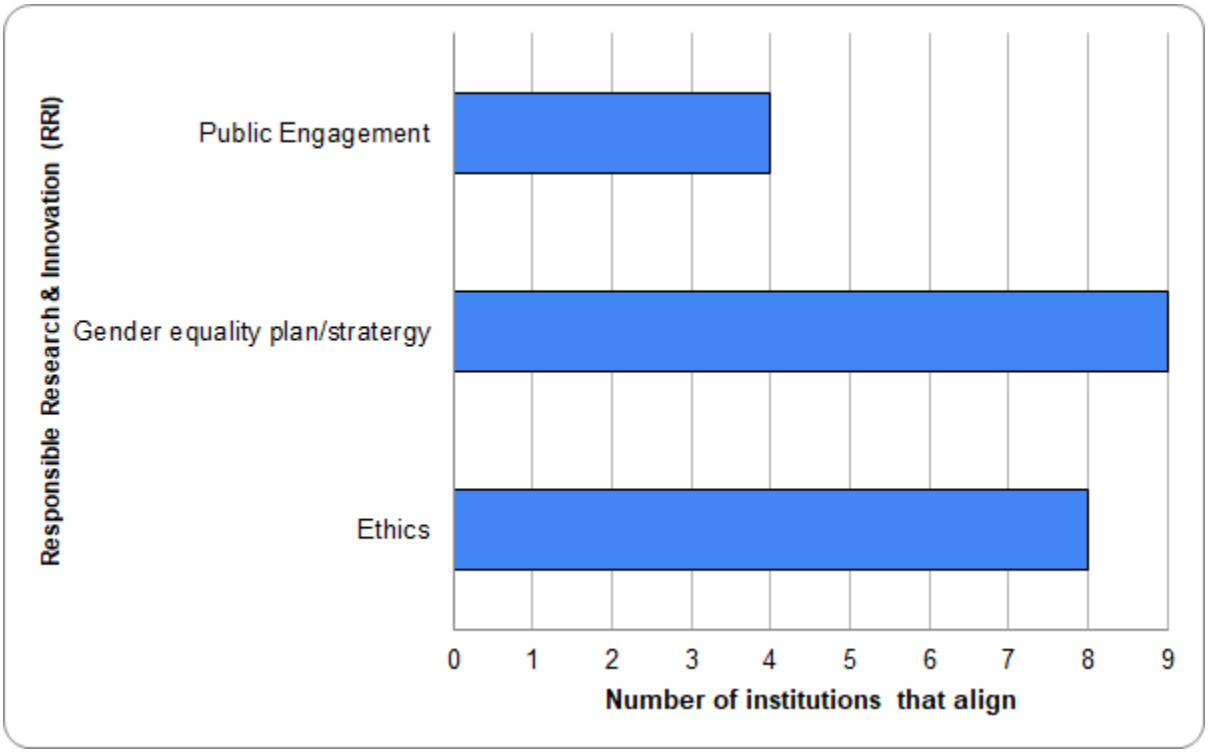


Figure 12. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Responsible Research and Innovation. Total number of institutions = 9.

Policy Area 8: Physical collections management

Nearly all the institutions align with Visitor access policy requirements. Just over half of institutions have a destructive sampling policy, whilst less than half for the protection of culturally sensitive objects.

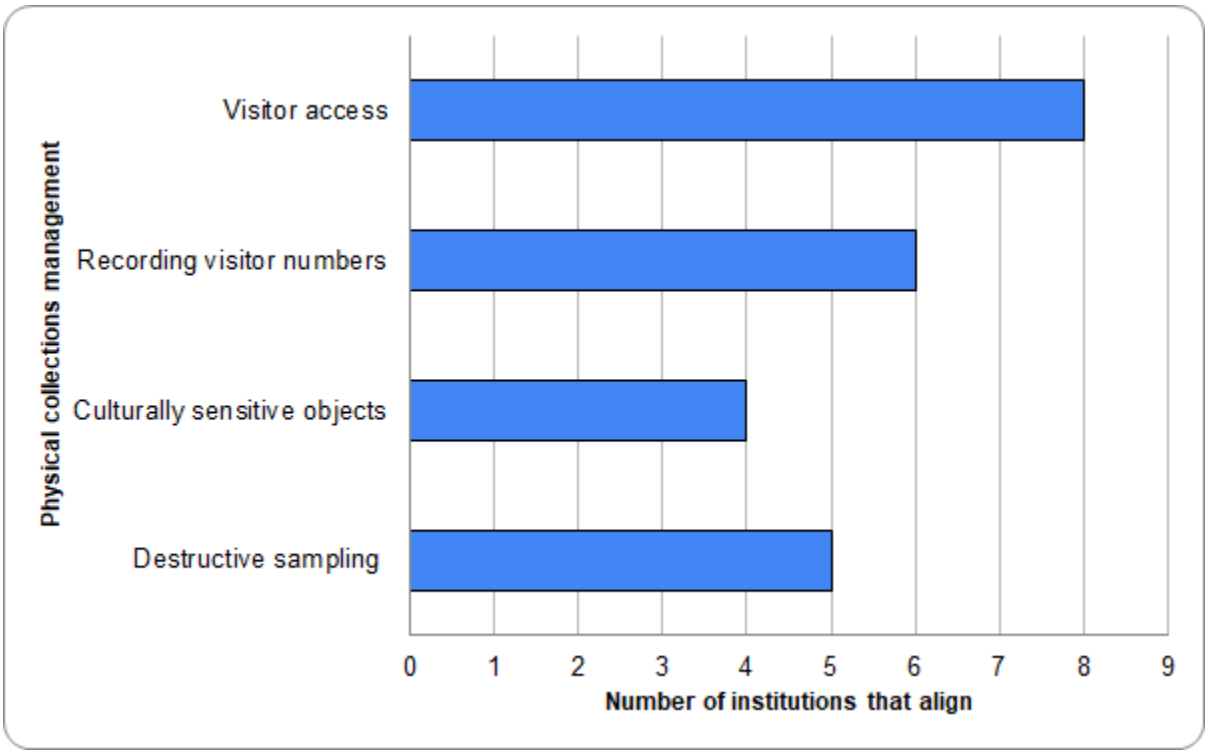


Figure 13. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Physical collections management. Total number of institutions = 9.

Policy area 9: Digitisation strategy & prioritisation

Four institutions did not fully complete the information for this policy area so the alignment number shown in **Figure 14** may be an artefact of this. Five institutions indicate that that they have a digitisation strategy, 4 indicate that it is sharable but only 2 provided links to the documentation. Four institutions indicate that they mint there specimens with unique identifiers but this is not apparent in from information provided in the documentation.

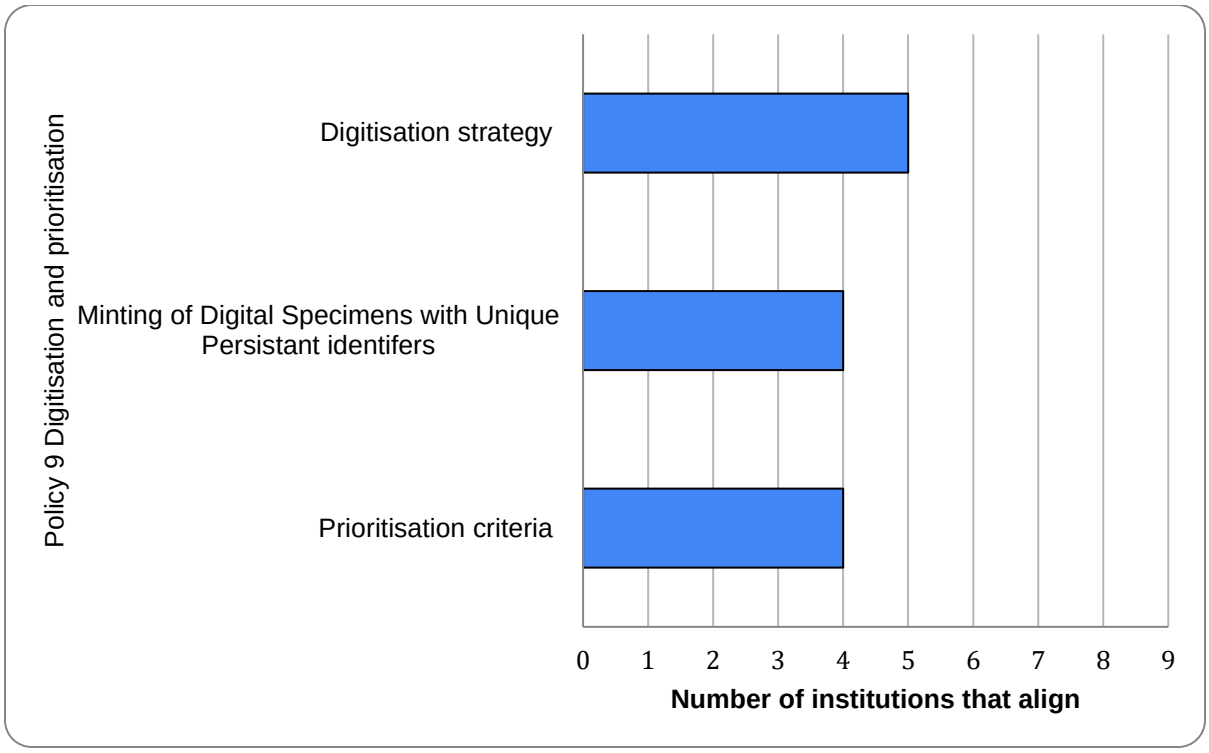


Figure 14. Bar chart showing the number of institutions that align with ELViS policy component requirements within the policy area Digitisation & prioritisation. Total number of institutions = 9.

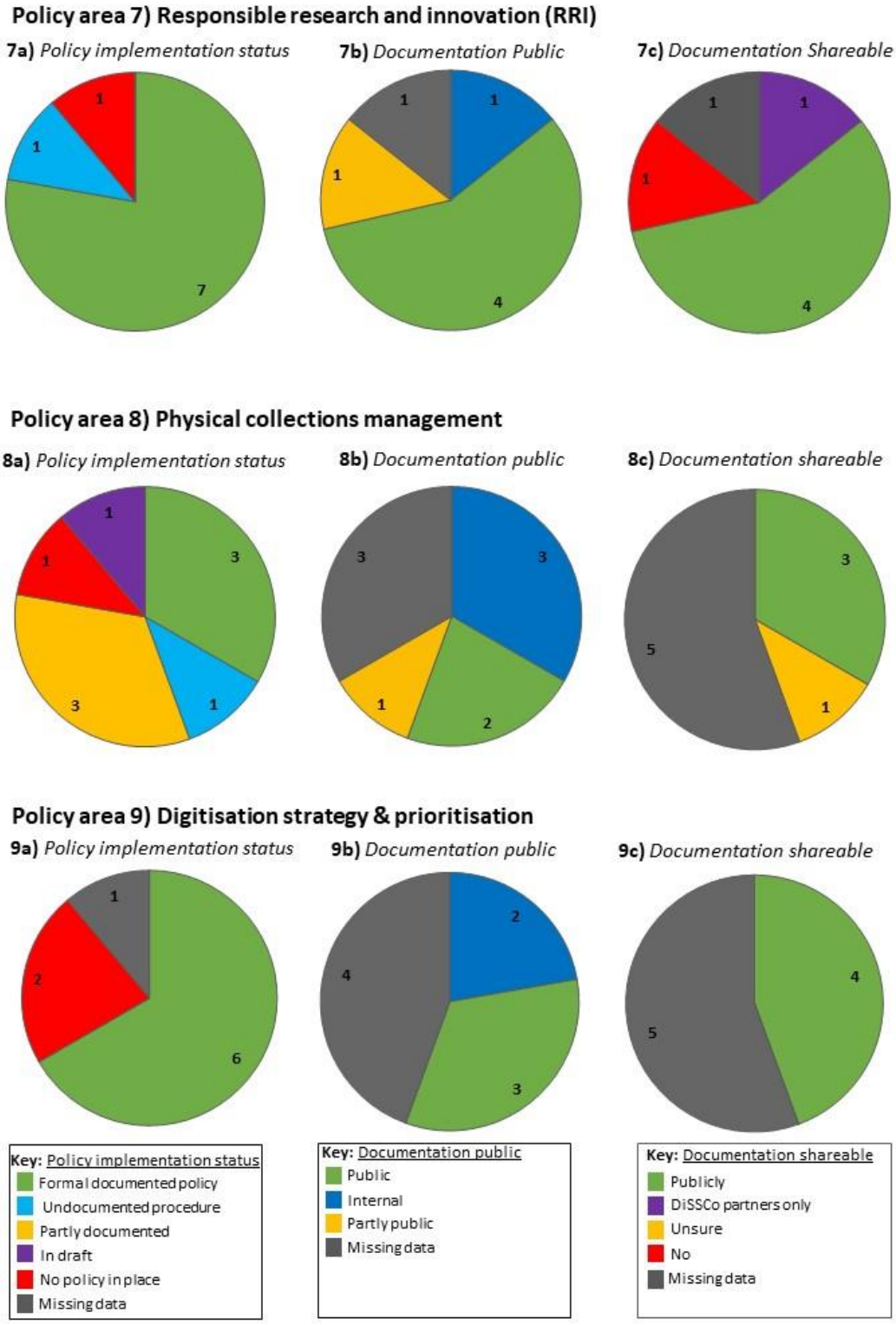


Figure 15. Pie charts presenting the Policy implementation status, Documentation Public and shareability for Policy areas: 7) Responsible Research and Innovation, 8) Physical collections management, 9) Digitisation strategy & prioritisation. Number inside the pie charts represent the number of institutions.

3.3 Discussion

Overall, the results of mapping the T2.1 partners policy data to the metadata categories, Policy Component, and Institution Policy Area, indicate that the metadata schema works for presenting DiSSCo policy needs and recording existing institution policies. However, feedback from partners said that there needs to be some improvement to the questions associated with the policy components, because some did not properly address the meaning of the policy component term and definition. The results generally show that there are no real boundaries, such as institution or national level policies, that prevent these nine institutions from aligning with ELViS policy needs; but this may not be the case for the rest of DiSSCo consortium.

The level of alignment is good for the components within the policy areas ‘Data and Digital Media sharing and publication’, ‘Collections access and Benefit sharing’, ‘Responsible Research Innovation (RRI)’, and ‘People data’. The likely reasons for good alignment in most of the components of these policy areas are that they contain requirements that are at EU level, for instance in the case of GDPR, Access and benefit sharing and components within RRI, which are required for joining EU projects. Another reason, in the case of policy areas: Loans management and Data and Digital Media sharing, some of the components are already at the core of institutions day to day operations, such as policies on visitor access, and loan policies, and data publishing to open repositories.

Despite some of the policy areas having good institution alignment, there is variability in alignment amongst the components within them. For instance, within policy area “Data and Digital media sharing” there is low alignment for the components referring to CC0 or CC-BY default licensing for research data, 3D images, and video and audio. The possible reasons provided by T2.1 partners for the slightly lower alignment are that some institution owners may require them to earn part of their budget themselves, and they may do this by selling high-end versions of such types of data for commercial use. With regards to research data there is still the traditional culture of publishing it in closed access journals, and the data is sometimes owned and actively used by the researchers that produced it. This may cause some limitations for using ELViS, and thus DiSSCo to enrich the digital objects with more complex data other than basic metadata, but it depends on the extent of the issues once other Institutions provide their alignment status in the Policy Self-Assessment tool. Increasing institution alignment for these components may encourage cultural change towards more open data and sharing at an earlier stage of research work. Another point to note about institutions showing an array of

default licences for specimens, such as NHMW, in these cases there is uncertainty on how these licences are applied to different specimens and to the extant (what % of what specimens have CC-BY or CC0 and what % do not. Why do specimens not have a more restricted licence?). Knowing the reason for the array of different licences may help gain insight in the extent of access and use limitations of specimen data. The alignment of FAIR data is not severely low with just over half of the institutions following, but there is no reason given why some institutions are not following FAIR principles.

The low alignment for overall for policy area data standards, as well as individual components such as ‘Link images to IIIF compatible servers’ (in Data and digital media sharing and publication), Identifiers for person authentication (in Personal data) is likely due to the fact that they are new concepts or still development so institutions have not had the time or resources to integrate them as policy yet. For example, for the policy components within the policy area data standards such as the use of Unique identifiers (for collection, specimen and institutions), MIDS, TDWG collections and the SYNTHESYS+ collection classification scheme is that they are new concepts or processes and therefore not widespread, and in the case of MIDS and TDWG collection description they are still under development. Work will need to be done towards engaging institutions in the use of these data standards, and how to support them in implementing them to the level of policy alignment needed. This low alignment for the components in this area may limit the quality of data needed by ELViS during its early stages and may initially hamper discovery and access to institutional holdings. For the low alignment of the data management plan within the policy area ‘Data Management and Curation’ is also likely to be due to being a new need and based on the findings of (Saarenmaa et al., (2019) with regards to the implementation open access data guidelines found that institutions are only found at the project level currently few have on at institutional level. It is unclear why most institutions currently do not align with the having a defined incident alignment process from feedback. This may be a problem from the point of view of operating ELViS' for transparency reasons and assurance to prevent data breaches that could impact ELViS and DiSSCo itself; data is removed and changed by people without authority to do so, as well as the insight into the level of service institutions can provide.

With regards to the results of analysing the T2.1 institutions information about the status of their existing policies for each area, (Metadata category: ‘Institution Policy Area’), it has been determined that for most of the policy areas, it is not necessary for the institutions to have formal

documentation. For instance, most institutions do comply with providing data to open access repositories without having policy documentation. With this point in mind, it is suggested that the phasing of questions posed in the policy self-assessment tool should be changed to 'Are you able or willing to….?' Rather than do you have….? A few institutions do not have their own institutional policy on GDPR because they see there is no need to document it since it is documented at the EU level. The benefit of having policies written in formal documentation is that there are solid reference guidelines that help the organisation in consistent decision making. However, the feedback from T2.1 partners also suggest that making it mandatory for institutions to document all the necessary policy components regarding DiSSCo could be a heavy administrative burden and use of resources. Even so, from the point of view of a ELViS user having a documented data management plan, visitor access policies, destructive sampling policies may be important for deciding which collection they want access to and planning research/work. The main reason for asking institutions for their policy documentation in the Policy Self-Assessment tool is for potential references to help other institutions who may not have anything in place for a policy area.

There are no policy documentation examples from the sample in which all components for a policy area are present; rather the available policy documentation provides fragmented elements of components as examples that can be uploaded to the DiSSCo knowledge base. Reference material that can be shared for the following policies:

- Data publication: Meise DMP, RBGK and CSIC
- Personal Data: CSIC, RBINS and RBGK
- Data curation and management: BGM
- Research and Responsible Innovation: CSIC, RBINS and RBGK.
- Digitisation and Prioritisation: RBGK and NHMW.
- Access and Benefit sharing: CETAF Code of Conduct.

4. Conclusions and future recommendations

SYNTHESYS+ Task 2.1, in collaboration with NHM London (leaders of DiSSCo Prepare 7.3) have delivered and tested a pilot policy alignment criterion in the form of a metadata schema. The policy metadata schema essentially provides a structured framework for describing and

cataloguing 1) DiSSCo service workflows (components) and their policy requirements 2) Institution policies (what institutions have and status of documentation) and how they meet the requirements of the DiSSCo service. Furthermore, it underpins a Policy Self-Assessment tool for DiSSCo partners. For now, the schema only presents the Service ELViS and its related policy requirements due to the fact it is the most significantly developed service and some of the policies are known or can be interpolated. However, ELViS was still under construction throughout this task, so some policy requirements may be missing due to uncertainty at this point, and those present may need to be further evaluated in the next phase of DiSSCo.

The current version of the metadata schema is here **Metadata schema V.2.** It will be uploaded to the central DiSSCo GitHub repository where it can be further expanded, improved and issues can be flagged.

The metadata schema was successfully tested by the T2.1 partner institutions who were asked to map their own institution policies to it, to assess the practicality, as well as seeing how they align with the identified policy requirements. There were no major practical difficulties in institutions comprehension of Policy areas, and their ability to map their policies. Analysis of the policy information that they provided indicates that there are no major blockages preventing institutions from complying with ELViS requirements. However, this may not be the case once the requirements are rolled out to all DiSSCo members.

Despite this achievement the metadata schema is not a finalised product, but rather one that needs to be improved, maintained, and expanded with other DiSSCo service policy needs. Throughout the process of creating the schema and conducting the policy mapping exercise, lessons have been learnt about how to approach the policy alignment for DiSSCo, as well as possible challenges identified for policy alignment. The following section summarises the challenges, future improvements identified through this work and provides recommendations for next steps. Recommendations are also given for ensuring the sustainability of the metadata schema, beyond SYNTHESYS+, and its expansion for DiSSCo services and other natural science related policy needs.

4.1 Current metadata content alterations

In task T2.1 partners have made significant effort in trying to identify ELViS policy alignment needs using available resources and defining them clearly within the framework of the structure

of the policy metadata schema. However, due to time constraints, the limited expertise in the group and constant new information as DiSSCo evolves the following alterations are recommended for policy components and the policy questions.

Recommendation 1: *Due to lately received information that DiSSCo will be responsible for minting digital objects rather than the institutions themselves. The name and definition of the policy component '**Unique persistent identifier for specimens (non-digitised and digitised)**' (with policy area Data standards) should be altered to just apply to physical specimen identifiers, and any of the related policy options that refer to Digital object identifiers (NSIDs, DOIs) should be removed.*

Recommendation 2: *The policy component 'Minting of Digital Specimens with Unique persistent identifiers' (in Policy Area Digitisation Strategy & prioritisation) should be removed based on the reasoning presented in Recommendation 1.*

Recommendation 3: *the term 'Link images to IIIF compatible servers' should be altered to reflect more of a policy than a technical requirement.*

Recommendation 4: *MIDS should be removed as Policy Component Options of the Policy Component Specimen data description standard, and instead defined as a policy component. A suggested name for the policy component 'Specimen Digitisation Level defined using Minimum Information about a Digitised Specimen''*

Recommendation 5: *The question relating to the policy component 'Incoming and outgoing loans' needs to be revised. A suggestion: Does your institution record incoming and outgoing loans?*

4.2 Facilitating institution alignment

The definition of what constitutes a policy in the context of DiSSCo has been somewhat vague; it remains unclear whether, for example, it only comprises guiding principles or all necessary operations or protocols (often specifically adjusted to e.g. national legislation, the institution's owner's wishes, other local circumstances). From the results of the mapping of institutional policies to the metadata schema it was concluded that for most of the policy areas and policy components needed for using ELViS, it is not necessary that institutions have formal policy

documentation if they can or are willing to align/fulfil. The benefit of having a policy as formal documentation is that it can be used as references to help other institutions to align their policies (if publicly available).

Results from the mapping exercise indicated that institution alignments are higher for the policy components that are legal requirements at EU level, which is expected since there is a stronger incentive if it is a legal requirement. Low alignment occurs for policy components in that new concept or still under development as is the case for most of the components under Data standards, Digitisation Strategy and Prioritisation, which are new and additional to the more traditional core policies of an individual institution. In some cases, for example the application of data standards, Institutions may have decided that a formal policy is not required as the institution is happy to adapt its practices to widely supported community standards

There were few policy documentations that could be used as references for other institutions. None of the documentation for the policy areas reference all policy components but rather provide fragmented elements. The policies documents that have permission will be provided to the DiSSCo knowledge base.

Recommendation 6: *Considering that formal documentation is not needed for institutions to align with most policy components, it may be helpful to indicate this in some form. For example, if a policy component does need to be in documented form, then this can be encapsulated in the policy component name or definition. Additionally, in the Policy Self-Assessment tool user interface there could be an explanation on why DiSSCo is asking for institutions policy implementation status. Alternatively, if the DiSSCo CSO decide they want to use the policy tool, to ask institutions to sign-up and confirm compliance for a particular service, rather than understand the policy landscape the questions posed to DiSSCo members could be rephrased from "Do you have a policy on…" to 'Are you able or willing to provide [whatever the policy requirement may be]?" However, this may cause alternations in how policy options are structured in the metadata schema. It should be carefully assessed whether formal documentation is really needed for certain policy areas/components, to ensure there is no unnecessary administrative burden. This accommodates smaller institutions, and institutions with tight budgets, that want to keep administrative work at a minimum.*

Recommendation 7: *DiSSCo should provide support and encourage institutions in aligning with the policy components that are new concepts and not strict legal requirements by being*

clear about what is required and why, which will allow institutions to consider policy change considering the advantages of the proposed service. Even if the institution does have policies which currently preclude the institution from adopting the service, the institution may consider updating or revaluating its policy considering the advantages that a service might offer to the institution. Furthermore, clear business cases for the need for DiSSCo services will help drive institutional engagement in DISSCO and thus the revision of policies which support involvement in DiSSCo. However as for technical services the need for these must be clearly articulated and benefits made clear.

Recommendation 8: *Policy alignment around the handling of sensitive data is low. This is an area where DiSSCo could take a lead to establish common guidelines to assist alignment in areas such as restriction of threatened species data or for culturally sensitive items for the DiSSCo network taking into account other initiatives such as the IUCN sensitive data access policy of the IUCN Red LIst for Threatened species data and Current Best Practices for Generalising Sensitive Species Occurrence Data from GBIF for threatened species or The CARE Principles for Indigenous Data Governance for culturally sensitive data.*

Recommendation 9: *To strengthen the provision of reference material for policy development, DiSSCo should look to CETAF for support, since the community has developed best practice guides for some policy areas such as the CETAF Code of Conduct and Best Practice for Access and Benefit-Sharing. In addition, within CETAF there are working groups that focus on Policy compliance (e.g., the policies and legislation working group). Guidelines for publication best practices are currently being developed under the CETAF e-publishing group for the DiSSCo linked EU project Biodiversity Community Integrated Knowledge Library (BiCIKL). Also, the Earth Science Group have done some research on EU country policies regarding the importing and exporting of cultural goods.*

4.3 Future expansion and developments

A major challenge of constructing the policy content within the metadata schema was the identification of the actual policy requirements for using ELViS. More specifically, although ELViS is the most developed DiSSCo-service during the timespan of this task, T2.1 had to decipher most policies from multiple documents, and some of the mentioned policy requirements were only described generally and have not been formally decided/endorsed. These are the main reasons for the long duration until constructing the content of the Metadata

Schema was completed. For the future inclusion of other DiSSCo services and the policy alignment needs, DiSSCo should streamline the process of identifying policy requirements.

As mentioned above, the metadata schema needs to be expanded to incorporate the policy requirements of other services such as the CDD - Collection Discovery Dashboard, or the SDR - Specimen Data Refinery, Knowledge Base, Helpdesk, AAI etc. (please refer to the community e-services brochure for a list of core services). Since ELViS as a core service follows the requirements of the DiSSCo architecture, we believe that the policy metadata schema produced in this covers the bulk of the policy areas that relate to other technical core services as outlined in the recently published. However, each of the other services may have additional and different policy aspects. In broad terms foreseen new additional policy Categories maybe Financial and Human resources.

Recommendation 10: *The process of identifying policy requirements DiSSCo services needs to be streamlined via a centralised process, to reduce the time and effort exhibited in this task. More specifically, DiSSCo service coordinators/providers, and/or the CSO should contribute to identifying the service policy needs and this could be done via a survey structured in the format of the metadata schema (e.g., providing a breakdown of the services/or elements of DiSSCo and mapping their policy needs. The fine-tuning of the terminology/definitions can be done by a dedicated policy task force made up of experts in the community including expertise from both the bioinformatics and collection management (See recommendation 13).*

Recommendation 11*: Institutions may have to comply with legal obligations, other than GDPR. For example, obligations may come from national legislation, or insurance policies (commercially valuable collection objects). Such obligations may fit in the policy component "Protection of sensitive data" where at present only one (single) science-based policy for restricting access to data is mentioned. Apart from "protection of endangered/vulnerable species" the acronym "etc." indicates other types of sensitive data. It seems necessary that DiSSCo, participating institutions and legal counsel discuss and decide what other sensitive data may be relevant and how to deal with them. ELViS may need a statement to describe formal governance.*

Recommendation 12: *In the DiSSCo Knowledgebase there should be a dedicated repository for the uploading of institution policies (that have agreed to be shared). This repository should have subsections labelled with policy areas in the metadata scheme, furthermore the dedicated*

repository should follow the structure of the properties set out in the metadata category Institution Policy Area.

4.4 Future sustainability and maintenance

As mentioned previously, the metadata schema is likely to continuously be amended and updated as new services are included, and as the DiSSCo policies become more concrete. Also, when the policy tool is rolled out to all DiSSCo partners there may be some areas highlighted for improvements. It is necessary for the metadata schema to be improved and updated because as mentioned previously the policy landscape is a dynamic one. For this it is important that there is some mechanism in place for new additions and alterations to be incorporated under a controlled authority or group of people consisting of different backgrounds, and constant review over a time at least say annually or but also if constant issues arise. CETAF as the community on which the DiSSCo infrastructure is built has the potential as the place where expertise from the community can gather.

Recommendation 13: *DiSSCo CSO and CETAF (as the network of natural science research institutions and their expertise), to collaboratively set-up a working group/task force with a chair/convener to set-up for the maintenance of the metadata schema. The group can consist of a core group of different expertise within CETAF and DiSSCo (publishing, collections, Digitisation, ABS, DiSSCo services etc.) that are able to dedicate adequate time. Other stakeholders should be welcome to provide suggestions for improvements. The group could meet on a regular basis or when necessary to review and improve.*

Recommendation 14: *The policy tool should have a way for users to flag any issues with the practicality of mapping or comprehension of content in the metadata schema. This may be done via the DiSSCo helpdesk. Under the SYNTHESYS+ T2.4 the helpdesk for ELViS is set up- but the helpdesk will be expanded to deal with other Services such as the Self-assessment Policy tool. A category for such requests can be added to the helpdesk (e.g., Self-assessment tool help, suggestions for improvement to policy schema). Any queries within the categories can be assessed by the helpdesk agent who can flag in the GitHub.*

Recommendation 15: *A workflow needs to be established to make sure that changes made to the metadata schema are updated in the policy tool. This may be automated or manual. For a*

manual update periodically and ad hoc, there is an authoritative person that communicates changes to those in charge of the policy tool.

4.4.1 Other user cases of the metadata schema.

There are some other use cases in closely related initiatives to DiSSCo. Having a wider use of the metadata schema may also help endorsement, maintenance, and improvement. CETAF presented the CETAF working groups (Legislation & Policy Group; Publishing group and/E-publishing group, Collections) with the metadata schema, first by inviting representatives to the MS24 workshop and by individual meeting. The purpose was to see how they envisage helping in the sustainability of the schema and if they see any further usefulness in the meta schema as a standard for presenting policy needs and evaluation relevant to aspects of the Natural Science Community.

The CETAF E-publishing group leader (Lawrence Benichou MNHN) provided a potential use case metadata schema, help the current work they are undertaking which includes monitoring and improving the FAIRness of data held in institutional taxonomic Journals; the monitoring of actions taken by the community to tackle CETAF strategy objectives, and the technical and political monitoring actions for diamond open access.

In addition, the E-publishing group is involved in the EU funded Biodiversity Community Integration Knowledge Library project (BiCIKL) - Task 6.3 integrating published narrative and Linked open Data. In brief the project aim is to build a new European starting community of Key research infrastructures in the domain of Biodiversity and establishing open science practices, providing new tools and methods for the harvesting, liberating, and linking of sub article data specimen, material citations, samples, sequences, taxonomic names, and treatments from literature. Task 6.3 involves the alignment of identifiers and bi-directional linking of FAIRified biodiversity data between literature and RIs (including publishing journal publishers to enable the distributed systems to link data. Their needs would fit into the metadata conceptual model as it is for them to express the needs for institutions to align and improve data fairness and linking; as well as see the status of institutions. A service could be added ‘Publication of Journals’ or the name of the specific publication platform, which could be further split into Service Components. It is likely that the policy needs mentioned such as licences and use of

journal identifiers could be integrated as new policy components into the Policy area Data standards and Data and Digital media sharing. The FAIR data policy component within the policy area “Data and digital media sharing” may need to be revised to better incorporate data from publications.

The Legislation and ABS group could use the metadata schema to help monitor the compliance of member institutions around ABS and to pinpoint areas of support.

Author Contributions

Contribution types drawn from: CRediT - Contributor Roles Taxonomy

Laura Tilley: Conceptualisation, Data Curation, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing. **Matt Woodburn:** Conceptualisation, Methodology. **Ana Casino:** Conceptualisation, Project Administration, Supervision. **Alan Paton:** Investigation, Writing - original draft, Writing - review & editing. **Edmund Schiller:** Investigation, Writing - original draft, Writing - review & editing. **Karin Wiltschke:** Investigation, Writing - original draft, Writing - review & editing. **Lisa French:** Investigation, Writing - original draft, Writing - review & editing. **Frederick Berger:** Investigation, Writing - review & editing. **Elspeth Haston:** Investigation, Writing - review & editing. **Laurence Bénichou:** Investigation. **Jonathan Blettery:** Investigation. **Gabor Csorba:** Investigation. **Patricia Mergen:** Investigation. **Carole Paleco:** Investigation. **Francisco Pando**: Investigation. **Patrick Semal:** Investigation. **Vince Smith:** Conceptualisation.

Reference List

Addink, W. and Hardisty, A. 2020. ‘openDS’ – Progress on the New Standard for Digital Specimens. *Biodiversity Information Science and Standards*. **4**.

Addink, W., Islam, S., Koumantaros, K. and Laskaris, N. 2020. *MS48 AAI Infrastructure Design* [Online]. Available from: https://osf.io/sc7vw.

Agosti, D. and Egloff, W. 2019. *Policy Component of ICEDIG Project Website* [Online]. Available from: https://zenodo.org/record/3366656#.YzLybUxByUk.

Bakker, H., Willemse, L., van Egmond, E., Casino, A., Gödderz, K. and Vermeersch, X. 2018. *Inventory of criteria for prioritisation of digitisation of collections focussed on scientific and societal needs* [Online]. Available from: https://know.dissco.eu/handle/item/79.

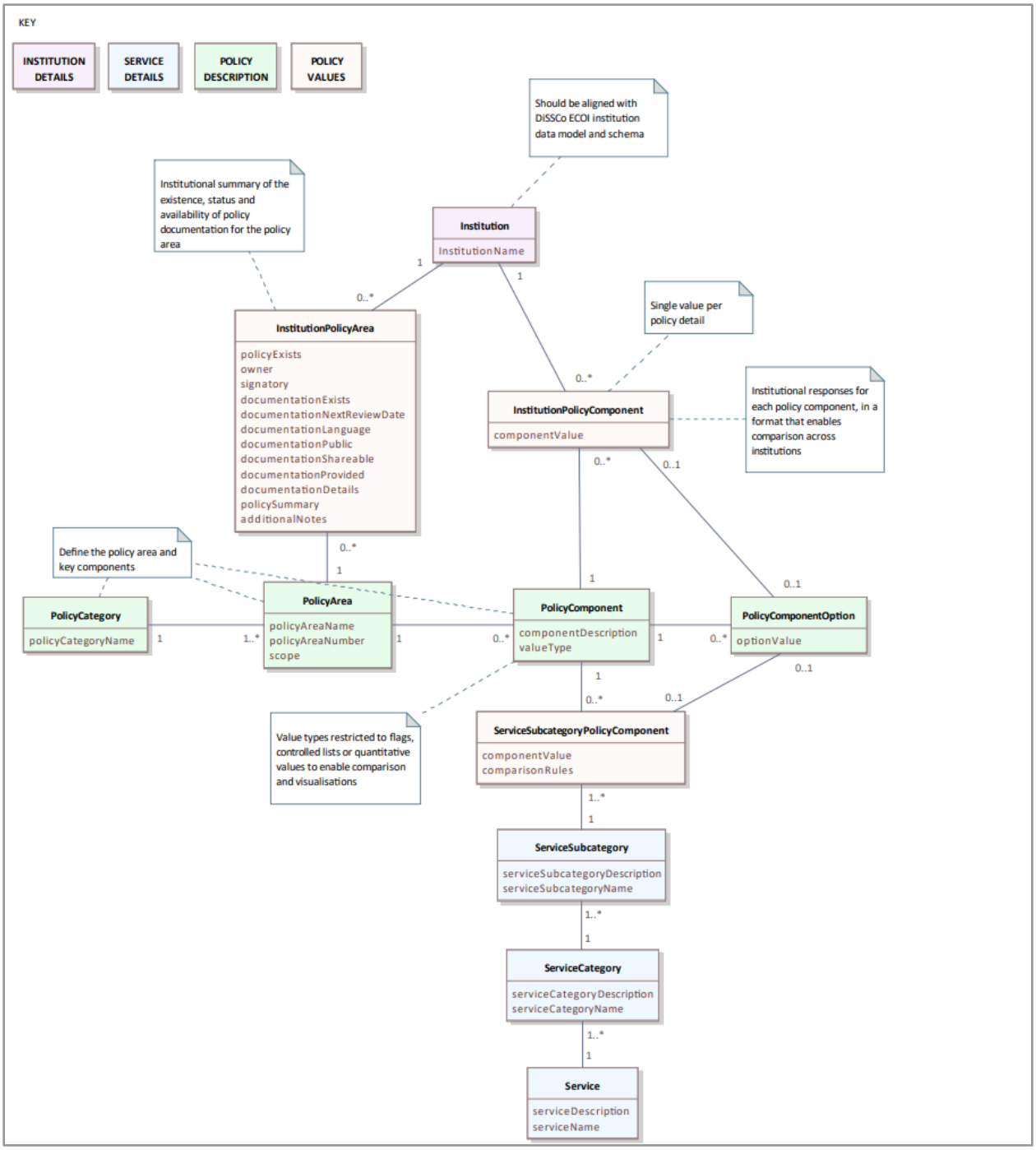
French, L., Woodburn, M., Blettery, J., Casino, A., Groom, Q., Hyvarin, M., Loo, T., Paleco, C., Pims, R.J., Scory, S., Semal, P., Tilley, L. and Smith, V. 2021. *MS - 7.5 Design of the*

policy framework tool [Online]. Available from: https://know.dissco.eu/handle/item/474.
Hardisty, A. 2019. *D6.6 Provisional Data Management Plan for DiSSCo infrastructure* [Online]. Available from: https://know.dissco.eu/handle/item/105.
Hardisty, A., Addink, W., Glöckler, F., Güntsch, A., Islam, S. and Weiland, C. 2021. A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). *Research Ideas and Outcomes*. **7**.
Hardisty, A., Saarenmaa, H., Casino, A., Dillen, M., Gödderz, K., Groom, Q., Hardy, H., Koureas, D., Nieva de la Hidalga, A., Paul, D., Runnel, V., Vermeersch, X., van Walsum, M. and Willemse, L. 2020. Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1. *Research Ideas and Outcomes*. **6**.
Hardy, H., Knapp, S., Allan, E.L., Berger, F., Dixey, K., Döme, B., Gagnier, P.-Y., Frank, J., Haston, E., Holstein, J., Kiel, S., Marschler, M., Mergen, P., Phillips, S., Rabinovich, R., Sanchez Chillón, B., Sorensen, M., Thines, M., Trekels, M., Vogt, R., Wilson, S. and Wiltschke-Schrotta, K. 2020. SYNTHESYS+ Virtual Access - Report on the Ideas Call (October to November 2019). *Research Ideas and Outcomes*. **6**.
Humphries, J., French, L., Woodburn, M., Tilley, L., Casino, A., Groom, Q., Hardy, H., Hyvärinen, M., Islam, S., Loo, T., Paleco, C., Semal, P., Scory, S., Worley, K. and Smith, V. 2022. *Assessment tools and direction map to the implementation of common DiSSCo policies - DiSSCo Prepare WP7 - Deliverable 7.3*.
Islam, S. 2022. *AAI landscape, ideas and plan* [Online]. Available from: https://docs.google.com/document/d/18S1WTJqROtx76a_eZax5Kc5NUPJywuroKoo87Nu-TeU/edit#.
Islam, S. and Addink, W. 2019a. *ELViS Design Plan* [Online]. Available from: https://docs.google.com/document/d/1sUWbcev46OqzgOLLip6tjBPJTNTy0P2ogDQ87vffuW8/edit.
Islam, S. and Addink, W. 2019b. *SYNTHESYS+ Data management plan, WP6 input* [Online]. Available from: https://docs.google.com/document/d/127CcFyce7uXhXNbdm9CioMBDYs89gNJbiFld2QLwpD0/edit.
Islam, S., Loo, T., Addink, W. and Hardy, H. 2020. *SYNTHESYS+ D6.1 Inventory analysis of dataflows* [Online]. Available from: https://docs.google.com/document/d/19KBAr3K17CDih3BTTs2bRVfU2oPJFAzYxqyd6A4pdME/edit.
Leeuwen van, C., Addink, W. and Van Dongen, W. 2019. *ELViS Project Plan* [Online]. Available from: https://docs.google.com/document/d/1jUbQAXYLtVX6VTBAKBg4IeNyyB60EXbYO--auDDksfc/edit.
Pain, E. 2017. To be a responsible researcher, reach out and listen. *Science*.
Saarenmaa, H., Agosti, D., Dillen, M., Egloff, W., Garnier, P.-Y., Groom, Q., Hardisty, A. and Raes, N. 2019. *Open access implementation guidelines for DiSSCo* [Online]. Available from: https://zenodo.org/record/3465285#.YuumvhxBw2x.
Teamwork 2015. Teamwork Crew Ltd.
Tilley, L. 2021. *Milestone 24: workshop to Integrate policies and produce harmonisation criteria* [Online]. Available from: https://docs.google.com/document/d/1fHteI4vBj_DiryySL5WL72pbnWAWA_AR/edit?usp=s

haring&ouid=112525402196466605612&rtpof=true&sd=true.
Tilley, L., Woodburn, M., Vincent, S., Casino, A., Addink, W., Berger, F., Bogaerts, A., De Smedt, S., Islam, S., Mergen, P., Nivart, A., Papp, B., Petersen, M., Santos, C., Schiller, E., Semal, P., Smith, V. and Wiltschke, K. 2019. *D2.2 Joint dashboard of collections asessment tools* [Online]. Available from: https://twk.pm/3ax5evp2mm.

Appendix

Appendix 1. First draft of the policy metadata schema conceptual model

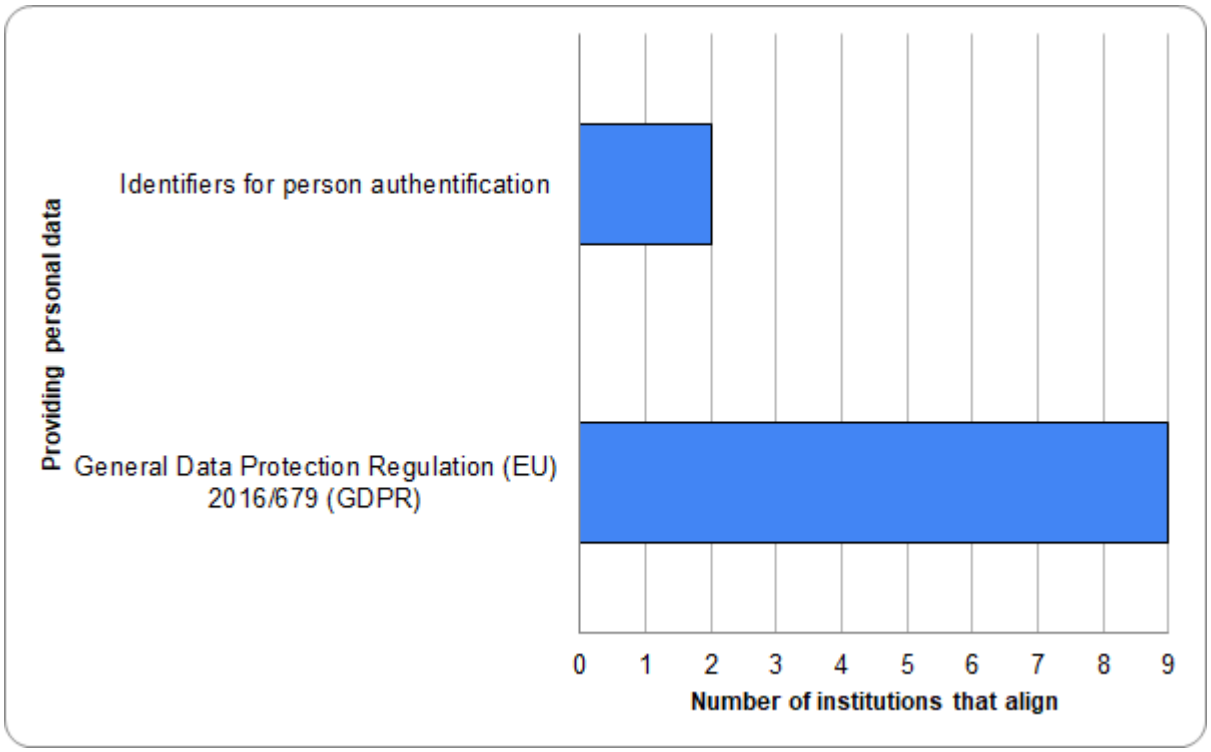


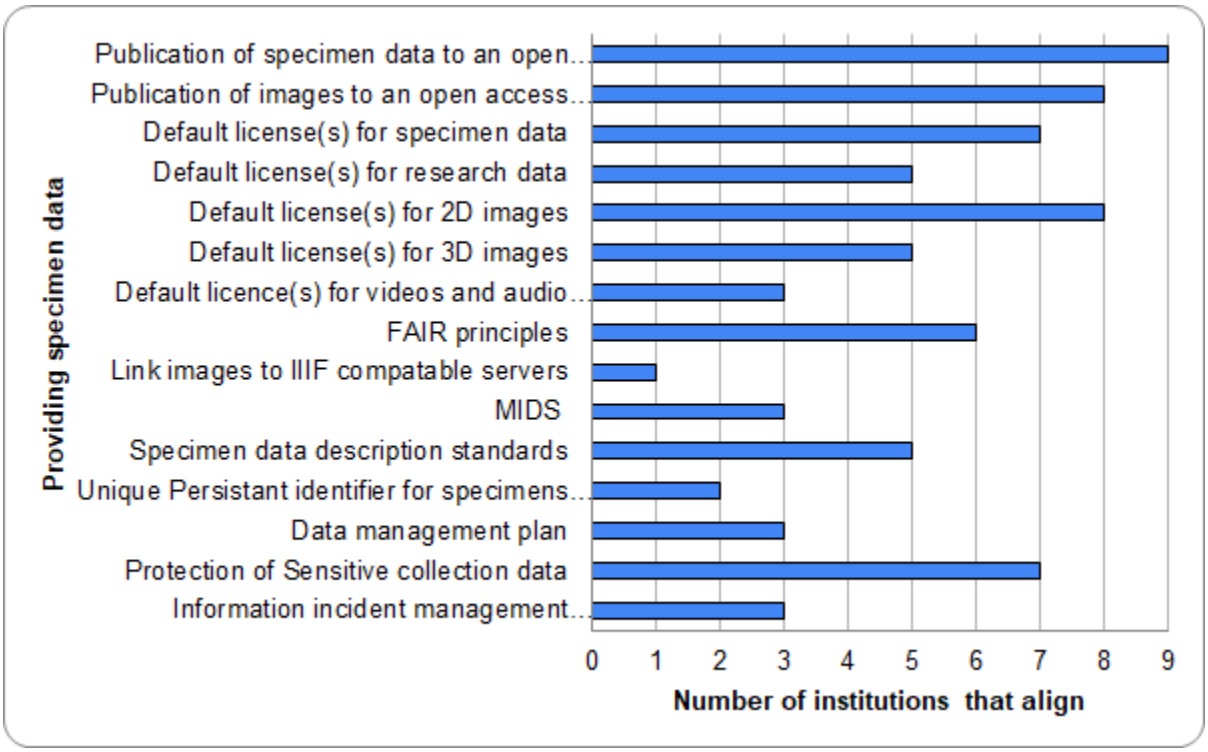
Appendix 2: Full links to metadata schema & case study data.

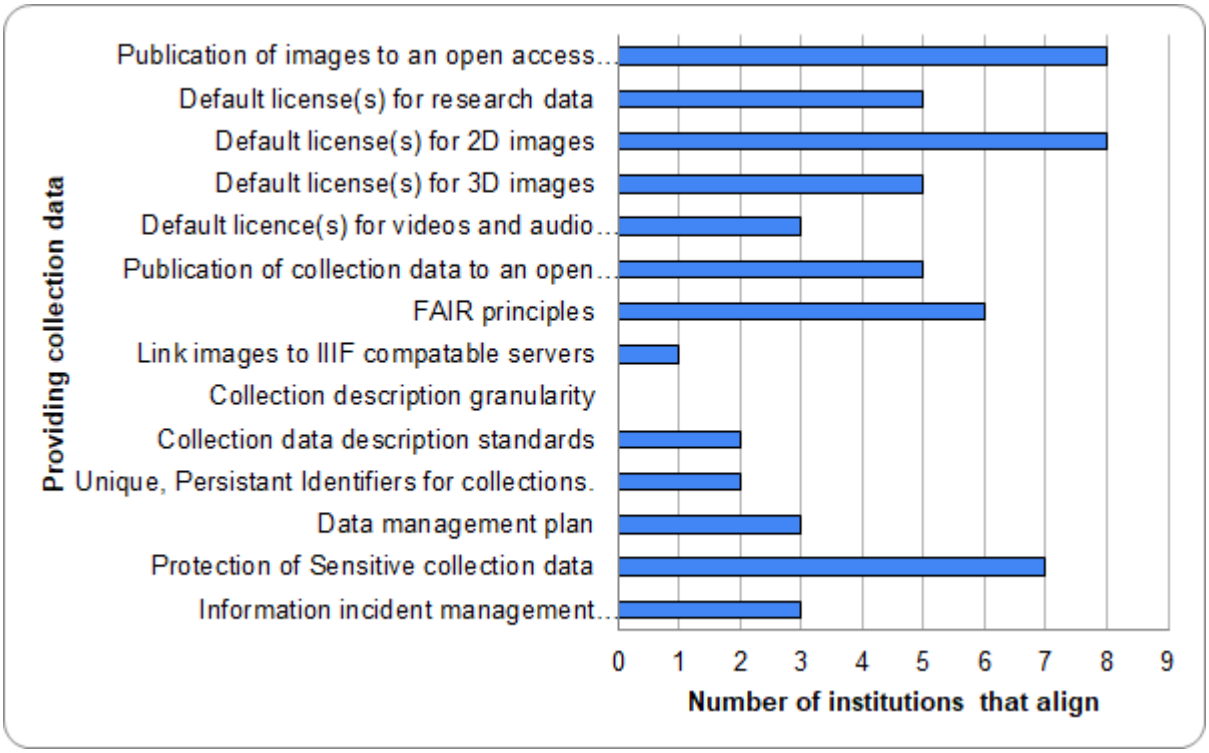
T2.1 Policy Metadata Schema V.2:
https://docs.google.com/spreadsheets/d/1H4B15urhCcIOTanMoyCHL_cD79aeA1-NqgR9VyU1hlk/edit#gid=1948292538

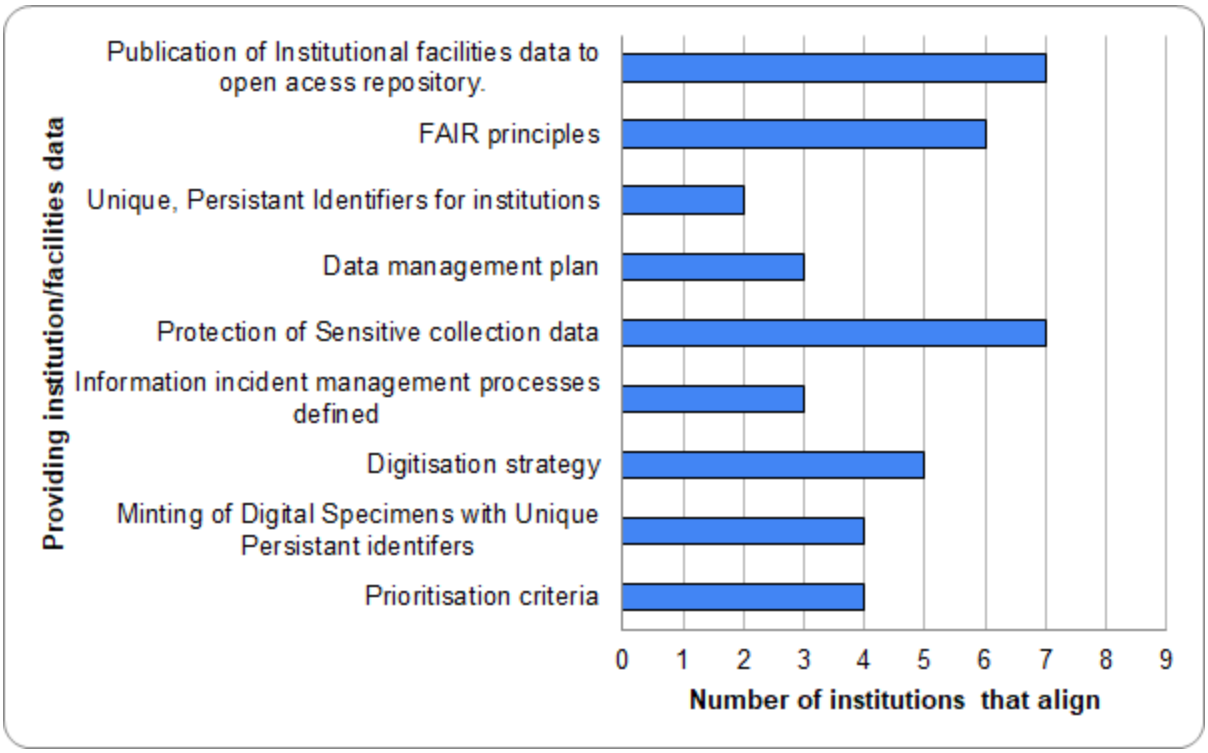
Mapping institution policies to meta schema V.2 (2022):
https://docs.google.com/spreadsheets/d/1REv9D0iIblIqyMMSBQOkz6U9dtBNN1RplECVcAZ_www/edit#gid=1640081436

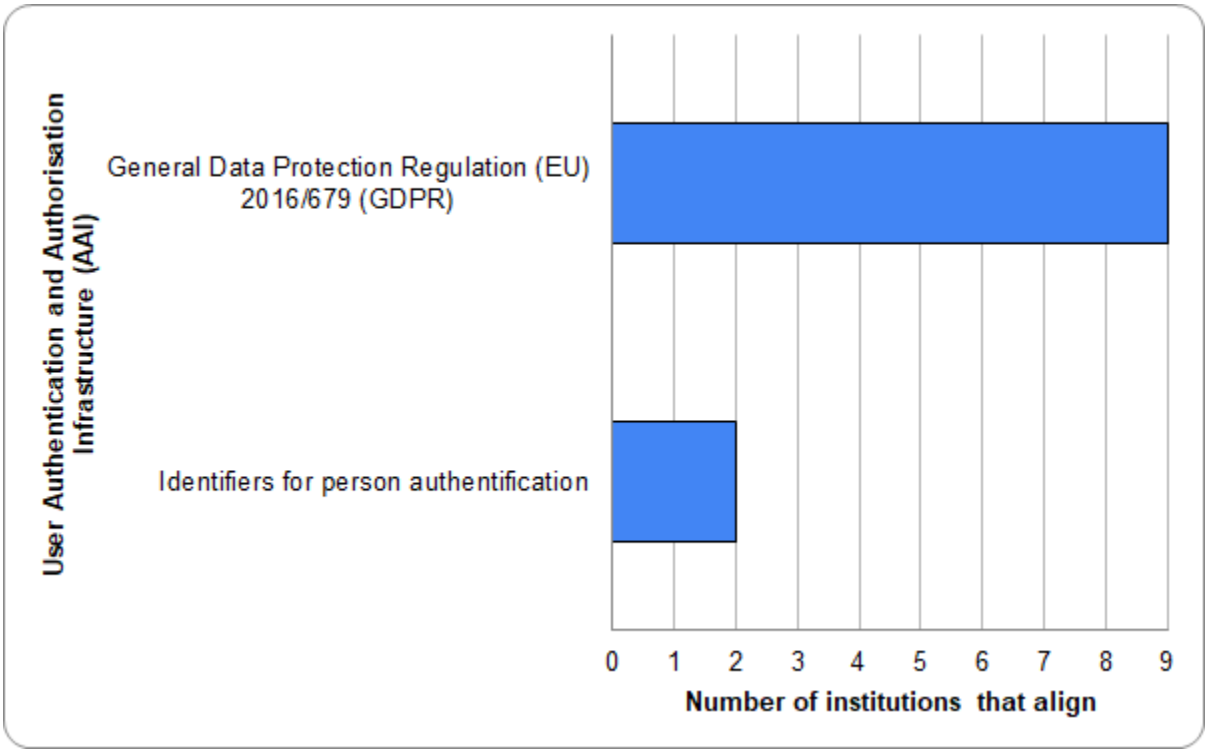
Appendix 3 Institution alignment of policy components per ELViS service component.

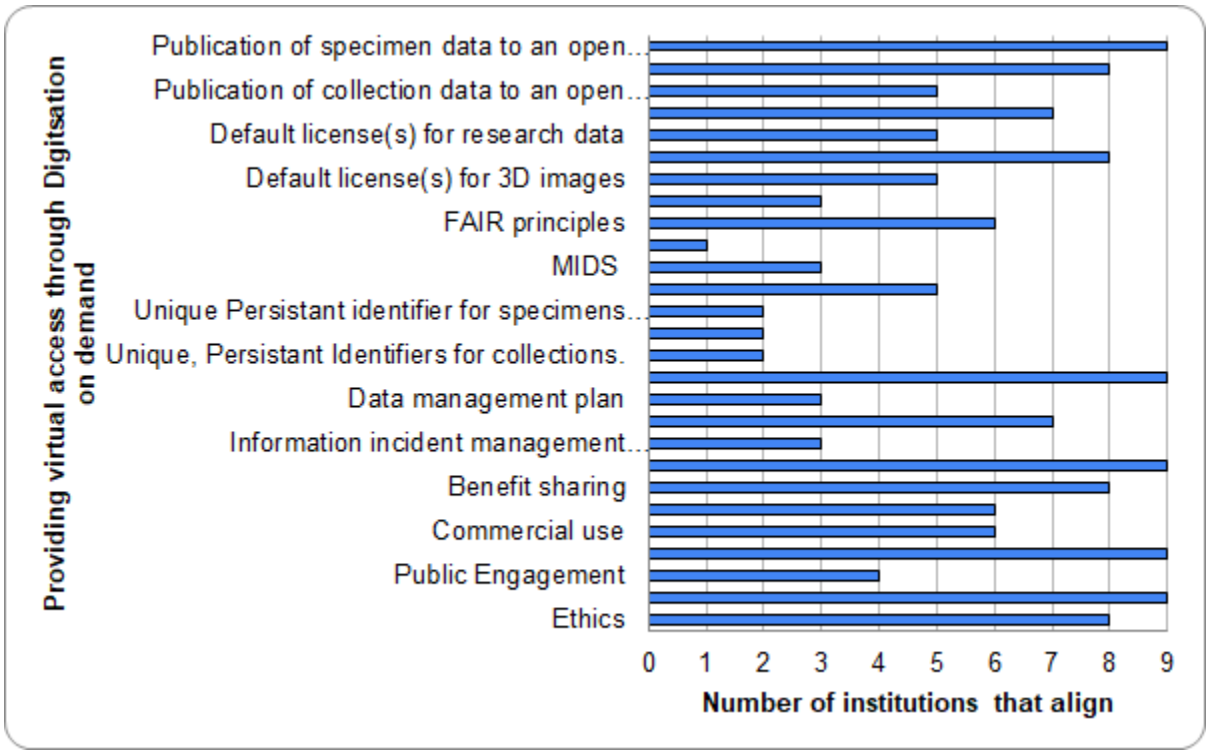


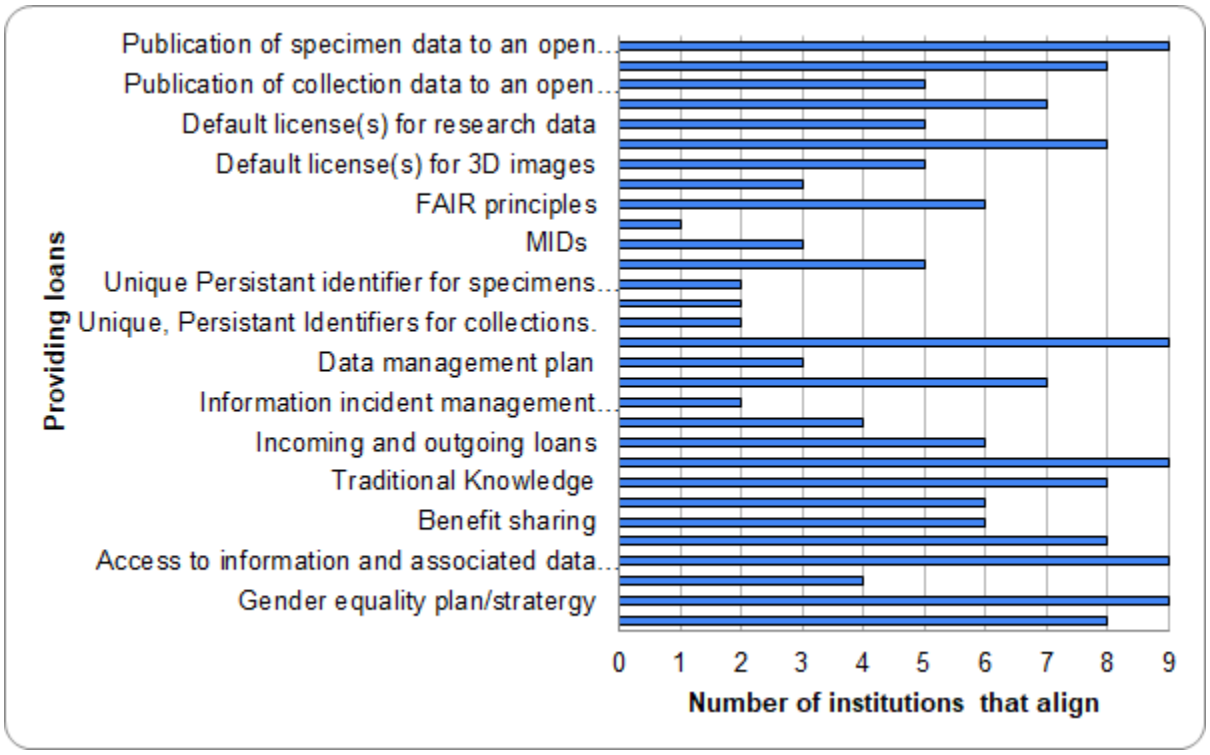

Providing specimen data
Publication of specimen data to an open...
Publication of images to an open access...
Default license(s) for specimen data
Default license(s) for research data
Default license(s) for 2D images
Default license(s) for 3D images
Default licence(s) for videos and audio...
FAIR principles
Link images to IIIF compatable servers
MIDS
Specimen data description standards
Unique Persistant identifier for specimens...
Data management plan
Protection of Sensitive collection data
Information incident management...
0 1 2 3 4 5 6 7 8 9
Number of institutions that align

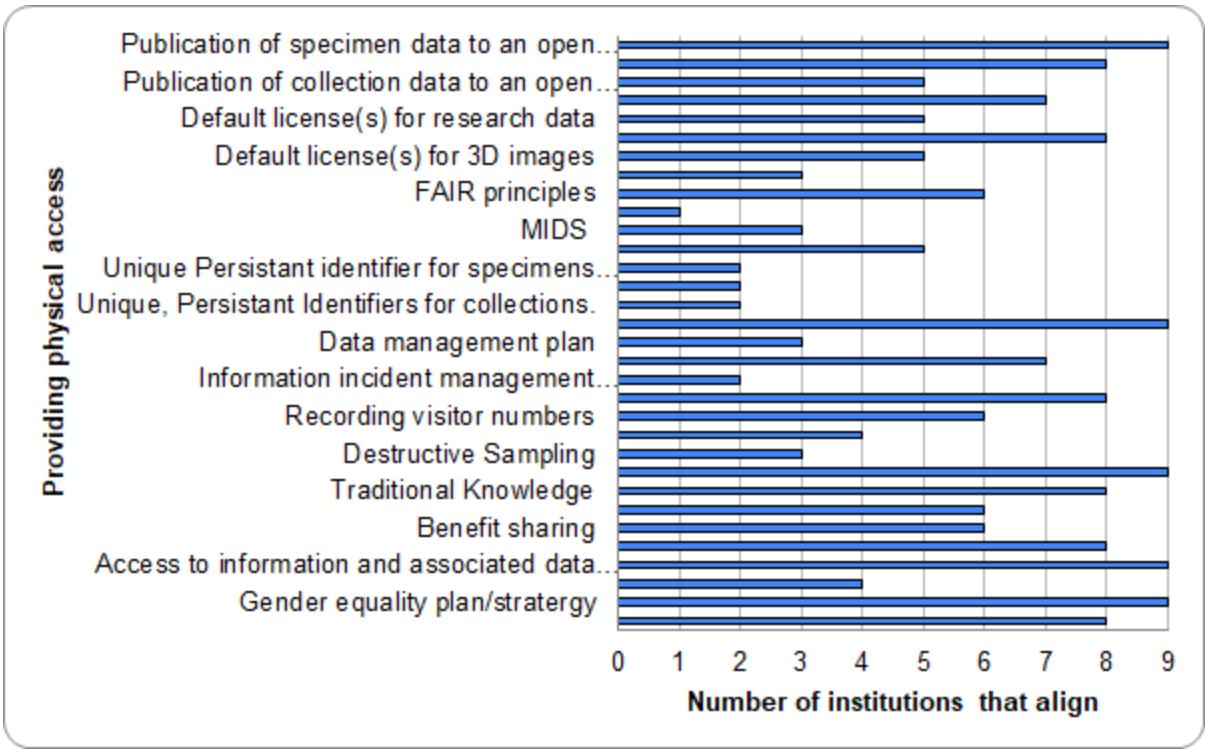

Providing collection data
Publication of images to an open access...
Default license(s) for research data
Default license(s) for 2D images
Default license(s) for 3D images
Default licence(s) for videos and audio...
Publication of collection data to an open...
FAIR principles
Link images to IIIF compatable servers
Collection description granularity
Collection data description standards
Unique, Persistant Identifiers for collections.
Data management plan
Protection of Sensitive collection data
Information incident management...
0 1 2 3 4 5 6 7 8 9
Number of institutions that align


Providing institution/facilities data
Publication of Institutional facilities data to open acess repository.
FAIR principles
Unique, Persistant Identifiers for institutions
Data management plan
Protection of Sensitive collection data
Information incident management processes defined
Digitisation strategy
Minting of Digital Specimens with Unique Persistant identifers
Prioritisation criteria
0 1 2 3 4 5 6 7 8 9
Number of institutions that align


User Authentication and Authorisation Infrastructure (AAI)
General Data Protection Regulation (EU) 2016/679 (GDPR)
Identifiers for person authentification
0 1 2 3 4 5 6 7 8 9
Number of institutions that align


Providing virtual access through Digitsation on demand
Publication of specimen data to an open...
Publication of collection data to an open...
Default license(s) for research data
Default license(s) for 3D images
FAIR principles
MIDS
Unique Persistant identifier for specimens...
Unique, Persistant Identifiers for collections.
Data management plan
Information incident management...
Benefit sharing
Commercial use
Public Engagement
Ethics
0 1 2 3 4 5 6 7 8 9
Number of institutions that align


Providing loans
Publication of specimen data to an open...
Publication of collection data to an open...
Default license(s) for research data
Default license(s) for 3D images
FAIR principles
MIDs
Unique Persistant identifier for specimens...
Unique, Persistant Identifiers for collections.
Data management plan
Information incident management...
Incoming and outgoing loans
Traditional Knowledge
Benefit sharing
Access to information and associated data...
Gender equality plan/stratergy
0 1 2 3 4 5 6 7 8 9
Number of institutions that align


Providing physical access
Publication of specimen data to an open...
Publication of collection data to an open...
Default license(s) for research data
Default license(s) for 3D images
FAIR principles
MIDS
Unique Persistant identifier for specimens...
Unique, Persistant Identifiers for collections.
Data management plan
Information incident management...
Recording visitor numbers
Destructive Sampling
Traditional Knowledge
Benefit sharing
Access to information and associated data...
Gender equality plan/stratergy
0 1 2 3 4 5 6 7 8 9
Number of institutions that align