DiSSCo related output

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

Title
WP3 Milestone 3.2 DiSSCo Digital Maturity Self -Assessment Tool - Design Blueprint

Author(s)	**Identifier of the author(s)**
Helen Hardy	Helen Hardy - 0000-0002-9206-8357
Anne Koivunen	Anne Koivunen - 0000-0002-3475-7971
Quentin Groom	Quentin Groom - 0000-0002-0596-5376
Patricia Mergen	Patricia Mergen - 0000-0003-2848-8231
Frederik Berger	Frederik Berger - 0000-0001-8400-3337
Rui Figueira	Rui Figueira - 0000-0002-8351-4028
Pedro Arsenio	Pedro Arsenio - 0000-0003-3860-9789
Alexandra Cartaxana	Alexandra Cartaxana - 0000-0001-8624-2198

Affiliation

Contributors

Helen Hardy - NHM London
Anne Koivunen - Luomus, Helsinki
Quentin Groom - Meise BG
Patricia Mergen - Meise BG
Frederik Berger - MfN Berlin
Rui Figueira - ULisboa
Pedro Arsenio - ULisboa
Alexandra Cartaxana - Ulisboa

Publisher

Identifier of the publisher

Resource ID

Publication year
2021

Related identifiers

Is it the first time you submit this outcome?
Yes

Creation date
30/04/2021

Version
1

Citation
DiSSCo Digital Maturity Self -Assessment Tool - Design Blueprint, H. Hardy et al, 2021

Abstract
This Milestone 3.2 report for DiSSCo Prepare Work Package 3 Task 3.1 sets out the initial design blueprint for a DiSSCo Digital Maturity Self-Assessment Tool, building on the analyses in the Milestone 3.1 report

'Improving Digital Capability - Case Studies and Analysis' (Hardy et al, Dec 2020) and in the Milestone 3.3. Report, including consideration of two existing tools in our sector. This tool is intended to support teams, institutions and national nodes in developing organisational readiness for provision of the DiSSCo services and data, helping them to identify and target areas for improvement. The aim is for this to tie in to future provision of training and support, as well as helping to identify the gaps at aggregate level where that training may be most useful. In addition , we believe there is a case for a platform that can support both this and the related Task 7.3 Policy Tool, such that these or other tools are consistent for users and can interact with one another where relevant, avoiding any duplication. This blueprint is intended for wider discussion among the DiSSCo members, so that tool content can be developed in more detail as part of the Deliverable for this Task.

Content keywords
technical

Project reference	**WP number**
DiSSCo Prepare (GA-871043)	WP3

Project output
Milestone report

Deliverable/milestone number
3.2

Dissemination level	**Rights**
Public	

License
CC0 1.0 Universal (CC0 1.0)

Resource type
Text

Format

Funding Programme
H2020-INFRADEV-2019-2

Contact email
H.Hardy@nhm.ac.uk

DiSSCo Capabilities – Additional Case Studies & Analysis

DiSSCo Prepare WP3 – Milestone 3.3

Work package Leader: Vincent Smith (NHM London)
Authors: Helen Hardy (NHM London), Anne Koivunen (Luomus), Quentin Groom (MeiseBG), Pieter Huybrechts (MeiseBG), Patricia Mergen (MeiseBG), Frederik Berger (MfN Berlin), Peter Giere (MfN Berlin), Rui Figueira (ULisboa), Pedro Arsenio (ULisboa), Alexandra Cartaxana (ULIsboa)

Abstract

This Milestone 3.3 report for DiSSCo Prepare Work Package 3 Task 3.1 is intended to update and be read alongside the Milestone 3.1 report 'Improving Digital Capability - Case Studies and Analysis' (Hardy et al, Dec 2020). It builds on the previous report by addressing additional case study examples of competency frameworks; carrying out further work to examine public sources of competency data about individuals in line with the next steps previously identified; and examining two digital maturity self-assessment tools to draw out learning that has informed Milestone 3.2 'Blueprint for a digital maturity self-assessment tool'.

Key words

DIGITAL MATURITY, CAPABILITY, COMPETENCY, LEADERSHIP

INDEX

01   INTRODUCTION

This report builds on the analysis in the Distributed System of Scientific Collections (DiSSCo) Prepare Work Package 3 Milestone 3.1 report 'Improving Digital Capability - Case Studies and Analysis' (Hardy et al, Dec 2020). That paper examined a wide range of examples of competency frameworks and related documents, however some additional examples were identified thereafter so this report discusses those and whether and how they provide relevant learning for DiSSCo.

Chapter 5 of the previous paper looked at ways to find capability data about individuals and organisations, identifying ORCID and LinkedIn as likely sources for further study. This report summarises additional work to examine these and other sources and their APIs, and to consider how effective these are likely to be in practice for finding colleagues with particular competencies or experience.

At the DiSSCo Prepare All Hands meeting in January 2021, discussion about the previous paper came to the conclusion that using Task 3.1 to develop a new, specific DiSSCo competency framework was unlikely to be productive at this stage, given the multiplicity of relevant frameworks already in existence but not being effectively used in our sector. It was agreed instead to work towards the development of a self-assessment tool for National Nodes, collections-holding institutions or teams to assess their digital maturity and set target areas for improvement, with an aspiration that this tool will be joined up with training and support as well as with other relevant tools or functionality such as the Policy Self-Assessment Tool envisaged in Task 7.3. The final section of this report therefore analyses two examples of existing tools or surveys relevant to digital maturity in the collections sector, and learnings from these have fed into Milestone 3.2 the Blueprint for a Digital Maturity Self-Assessment Tool, which will then be refined further moving towards the Task 3.1 deliverable in Summer 2021.

02 Additional competency framework examples

This Chapter builds on the competency framework examples analysed in the DiSSCo Prepare Milestone 3.1 report 'Improving Digital Capability - Case Studies and Analysis' (Hardy et al, Dec 2020). Additional examples were suggested to the group and these are reviewed below – these include examples of competencies and roles but also of aspects relating to digital maturity. Dedicated digital maturity tools are discussed in Chapter 4.

2.1 Belgian dictionary of competences[1]

According to the Belgian Federal Administration's dictionary of competences, "*Competence is the set of knowledge, skills, values and attitudes that the worker applies to function well in his function; in other words to achieve the results expected of him*". Concerning skills, these should describe HOW a function should be performed and not WHAT it consists of (the tasks it involves). The central assumption is that "*Different functions require a different set of skills*". These skills should then be measurable and developable. This approach relies on the definition of competency profiles, which contribute to:

- Identifying the skills necessary for the performance of a given function;
- Comparing the skill level of an employee with the skills defined for their specific function;
- Focusing on the development of skills that will further improve employee services.

This led to the development of competency profiles for a total of 25 'job families' within federal administrations. In these competence profiles, a distinction is made between 5 skill groups:
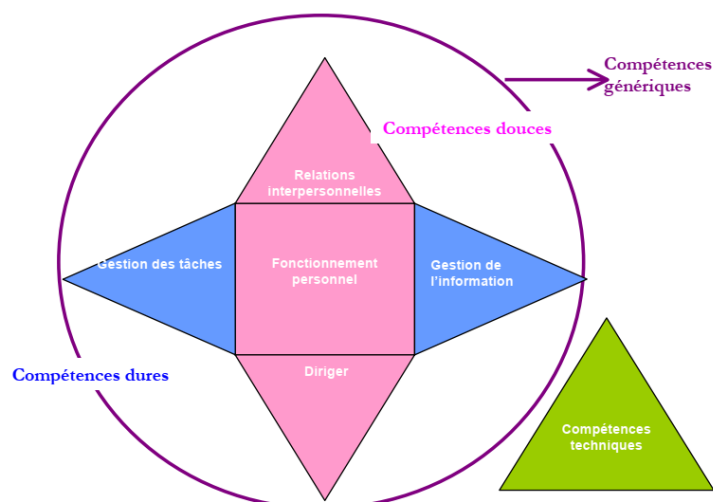
1. **Information management**: process information, facts, perceptions, knowledge and ideas for developing new information and knowledge;
2. **Task management**: organize and structure work, from organisational level to executive level, with a view to the optimal development of tasks and activities;
3. **Leadership**: manage and lead people in a hierarchical context;
4. **Interpersonal relationships**: managing relationships with others, excluding direct line management;
5. **Personal functioning**: management of own services and development.

A supplementary skills group is added to these, consisting of the **Technical skills** needed to carry out the function.

At a higher level of abstraction, these skill groups can also be defined as 'soft skills' (Leadership, Interpersonal relationships and Personal functioning); 'hard skills' (Information management and Task management); and 'Technical skills', as seen in the next Figure:

---

[1] https://fedweb.belgium.be/sites/default/files/BOSA%20Woordenboek-Dictionnaire-W%C3%B6rterbuch.pdf

At the organization level, skills management should be a specific component of the strategic policy of the organization, integrating HR processes to ensure that individual skills are optimally adapted to the objectives of the organization.

This framework has many similarities to those reviewed previously. The five skill groups are a useful way to summarise particular aspects in a way that cuts across job families.

2.2 IDigBio/CSIRO Biodiversity Informatics Manager Personal Skill Set[2]

This short document is focused on the observed profile of successful 'biodiversity informatics managers' (also described as 'Information Technology professionals'). It acknowledges the existence of differences regarding the 'capacity to influence' and dedicate time to these tasks according to the size and 'configuration' of the organization, stressing the importance of institutional support to biodiversity informatics managers, as a key feature in successful cases.

The following table illustrates some of these variations (Based on IDiGBIO Biodiversity Informatics Management Working Group, n.d.).

	Large institutions committed to mass digitization	Medium sized or decentralized organizations	Small institutions
Dedication regime	Exclusively dedicated, since digitization is at the core of the institution's mission	Serve several programs within a large biological sciences department	Serve several departments or colleges
Dedicated time	Pursue their role full time or nearly so	Divide their time among the different programs	(May develop their tasks in part time regime)

Upper level institutional support, including the existence of adequate funding, coordination of digitisation activities across the institution, adoption of institution-wide digitisation policies, the adoption of organizational structures allowing for rapid response to digitisation challenges and

---

[2] https://www.idigbio.org/wiki/images/f/fe/SkillSet.pdf

concerns, and standardisation of database management systems across the institution are also considered of the utmost importance.

The most successful biodiversity informatics managers observed by the IDigBio workgroup were:
- personable and positive;
- oriented toward service to staff;
- skilled in **personnel management**;
- oriented toward staff education and training;
- oriented toward rapid response to requests and suggestions;
- open to new and/or enhanced protocols;
- observant for methods to increase efficiency and productivity;
- **trained in the biological sciences**;
- experienced with and skilled in **collections management**;
- skilled in the **installation and use of technology**;
- skilled in the creation and customization of database interfaces and digitization software and equipment.

Often, the most successful professionals were biological scientists with combined interests in technology, biology, and collections management, since this mix of expertise allowed them to bridge the knowledge gap between collections managers and information technology standards, and to ensure the successful implementation of digitization equipment, workflows, and protocols.

This is very specific to a particular role, and also defines this role a little narrowly in relation to digitisation, however this role is highly relevant to DiSSCo and many of the observed success factors align with other competencies reviewed.

2.3 Atlas of Living Australia guidance - Digitisation: A strategic approach for natural history collections[3]

This report from the team from *Atlas of Living Australia* provides guidance to assist data providers in creating and managing digital content derived from their collections. It also guides institutions along the path of developing a digitisation program, by:
1. Explaining the role (and benefits) of digitisation in natural history collections;
2. Nominating key concepts and providing a model to access the institution's relative digitisation performance;
3. Explaining the main challenges an institution must overcome to implement a digitisation project; and
4. Providing guidance on the planning and implementation of a digitisation project.

An especially interesting element in this document is the 'digitisation maturity model', which allows for a quick diagnostic on where the institution stands regarding digitisation performance and level of maturity:

---

[3] https://www.ala.org.au/wp-content/uploads/2011/10/Digitisation-guide-120326.pdf

Specifically, every step of the digitisation plan is described, recommended practices are suggested and key questions are presented (see box below), as a way to establish where the institution is: 'where to aim' and 'how to get there'.

Key questions
- ❑ What standards are already in use in the institution?
- ❑ What new or revised standards are needed?
- ❑ What are the implications for your IT systems, work practices, etc of the chosen standards?
- ❑ What policy does it need?
- ❑ What processes will be used to implement the standards?

Finally, "*Annex 7: Suggested duties for a digital collections manager*" is also worth mentioning, since it specifies a number of duties for the digital collections manager, while stating that the holder of this position should be experienced in all aspects of digitisation and the management of digital assets, as well as knowledgeable in natural history collections or biodiversity.

2.4 The UK Collections Trust Collections Management Competency Framework[4]

The Collections Management Competency Framework from the UK Collections Trust charity defines the collections management skills and behaviours which a museum needs to develop, manage and sustain collections so that they can be used by the public. It is a high level framework which can be used as a starting point to define, plan and manage skills needed in a museum. The Collections Management Competency Framework defines 14 Areas of competency which are grouped into 4 clusters, as shown by the next figure.

---

[4] https://collectionstrust.org.uk/wp-content/uploads/2016/11/Collections_Management_Competency_Framework_2016_v2.pdf

The competencies in each of the four competency clusters are described in 3 levels, which are intended to reflect different stages in the learning and demonstration of competencies, starting with the ability to *Comprehend*, moving through the ability to *Apply*, and ending with the higher level competency, the ability to *Synthesize and evaluate*. Finally, it is assumed that a person demonstrating a competency at the highest level should already be demonstrating competency at the lower levels. This is similar to many of the other frameworks reviewed.

2.5 Digital skills for FAIR and open science[5]

The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science. Its development started in May 2015, when the European Commission proposed creating an Open Science Cloud for Europe to the Competitiveness Council. The aim was to federate existing research data infrastructures in Europe and realize a web of FAIR data and related services for science, making research data interoperable and machine actionable following the FAIR guiding principles.

The EOSC Skills and Training Working Group (WG) was formed in 2020 to identify a framework for building competence and capabilities for EOSC. This work was mentioned at section 3.5 of the previous DiSSCo Prepare Milestone 3.1 report, but has now reported. Their report includes policy and implementation recommendations for a range of stakeholders to facilitate advancement in this area. Especially relevant to mention are the definitions used:

- **Competence** or **competency**: An element (topic) of theory or practice e.g., 'workflow set-up and management', combined with an expertise level to indicate whether someone has an awareness of the area, or an ability to do it, or expert knowledge of it. **[individual level]**
- **Capability**: Competence applied at a research team or organisational level, with a defined level of expertise and responsibility, to perform a service role or work in the EOSC environment. **[organisational level]**
- **Skill**: A competence or capability acquired or applied in a specific context, e.g., producing a research output or deploying a service. A skill may be specified in a 'skills user story'. A badge

---

or certificate may provide evidence that a skill has been acquired, and a publication, personal profile, portfolio or CV may provide evidence that a skill has been applied.

The working group was also tasked with providing a framework for building competence and capabilities for EOSC in order to increase alignment on key skills and training components, as well as identify the skills and training structures needed to make EOSC sustainable, and determine how these could be embedded by different EOSC stakeholders.

A gap analysis was performed and the following issues were identified:
1) Lack of open science and data expertise;
2) Lack of a clear definition of digital professional profiles and career paths for these roles;
3) Fragmentation in learning and training resources.

Furthermore, a Framework of Actors in the EOSC ecosystem was prepared, in order to establish what exactly are the different skills and roles required to develop and use a complex federated infrastructure such as EOSC. The framework identifies ten roles, rather than job titles, as it is likely that in some cases the same person could undertake different roles (*e.g.*, trainer and researcher, data scientist and research software engineer, etc).

For each of the ten roles within the EOSC ecosystem, the report presents a description of the role, one situational example and a list of specific skills, along with a set of core transversal FAIR and open science skills. While these roles refer to Libraries and data, this could readily also be understood in a wider collections context. There is wider thinking within DiSSCo Prepare about the interaction with EOSC, and this model may well be relevant to further thinking about DiSSCo roles and capabilities.

03 Extracting competency information from public sources

3.1 Overview

Organisational capabilities depend to a large degree on the individuals working there. These capabilities are not always visible outside these organisations, and sometimes not even fully known within the institutes themselves. Various platforms are available where personnel may share information about their own individual skills and work experience. Prime examples are so-called professional social networks, such as LinkedIn and ResearchGate. LinkedIn is a more general platform, often used to connect people to vacancies or for professional networking. ResearchGate is more specialized as a unique online presence for published Researchers, where details of publications and projects can be shared. ORCID was set up to help disambiguate authors or scientific publications, however, ORCID profiles can contain information about an author's employment history. Other places where the activities, knowledge and skills of employees could be derived, are software repositories such as GitHub, where people can share their code and which are increasingly used for project management. Finally, public information about notable people is also available in general knowledge graphs such as Wikidata.

When looking for the skills present in institutions, these online repositories are possible sources to gather this information. Our goal was to test whether we can use the information freely available on these platforms to get an overview of the skills held by the employees in specific institutes or organisations. Ideally, we wanted to source information on everyone involved in collections, from technical staff to researchers, from curators to gardeners. Our expectation here is that different people and occupations use different platforms.

We also investigated whether we could enrich information that might be already at hand using open datasets or commercial APIs. This can be on a general level, connecting information with job titles, or specific to a person by querying commercial databases for any information they might hold about staff members.

Here specifically we attempted to retrieve from external sources: the people employed by a research institute; their job titles; and their skillsets and tools used, in order to facilitate the exchange of resources and to identify gaps in the capabilities of institutes.

3.2 Online public resources for competencies

3.2.1 LinkedIn

Many people maintain a LinkedIn profile. These profiles can contain information about past employment, skills, certification, education and more. The fact that LinkedIn is also used to fill vacancies is an incentive to maintain an accurate profile for those who are looking for employment. This incentive is perhaps less for those in a long-term engagement with a research institute. In the same vein, early career scientists are much more likely to add specific details, such as completed self-assessments for specific skills.

LinkedIn has a powerful API[6], but their privacy policy does not allow extraction of data related to experience, education, skills and certification, except with direct authorization from each LinkedIn profile.

3.2.2 ResearchGate

ResearchGate is a social networking website mainly aimed at scientists and researchers. Profiles are automatically created and disciplines assigned based on publications. Further information is self reported by the profile subject, and thus of high value. It contains information about publications, employment, grants, skills and used tools. However, it does not offer a (public) API which makes it difficult to assemble the data required.

3.2.3 Wikidata

Wikidata is addressable through the RDF query language SPARQL, which allows us to produce queries that are difficult to answer otherwise. One example would be to retrieve the job titles of everyone employed by an organization, such as a museum or a subclass of a museum; or for a list of research institutes, to retrieve all their staff members and their publications.

While the strength of the platform is the accumulation of information from different data sources and its openness, the main limitation for its use in compiling competencies is that the coverage of staff is not nearly complete enough. Furthermore, even for those who have a Wikidata page, interesting fields such as specialty and publications are either not or insufficiently populated.

---

[6] https://docs.microsoft.com/en-us/linkedin/shared/integrations/people/profile-api?context=linkedin/consumer/context
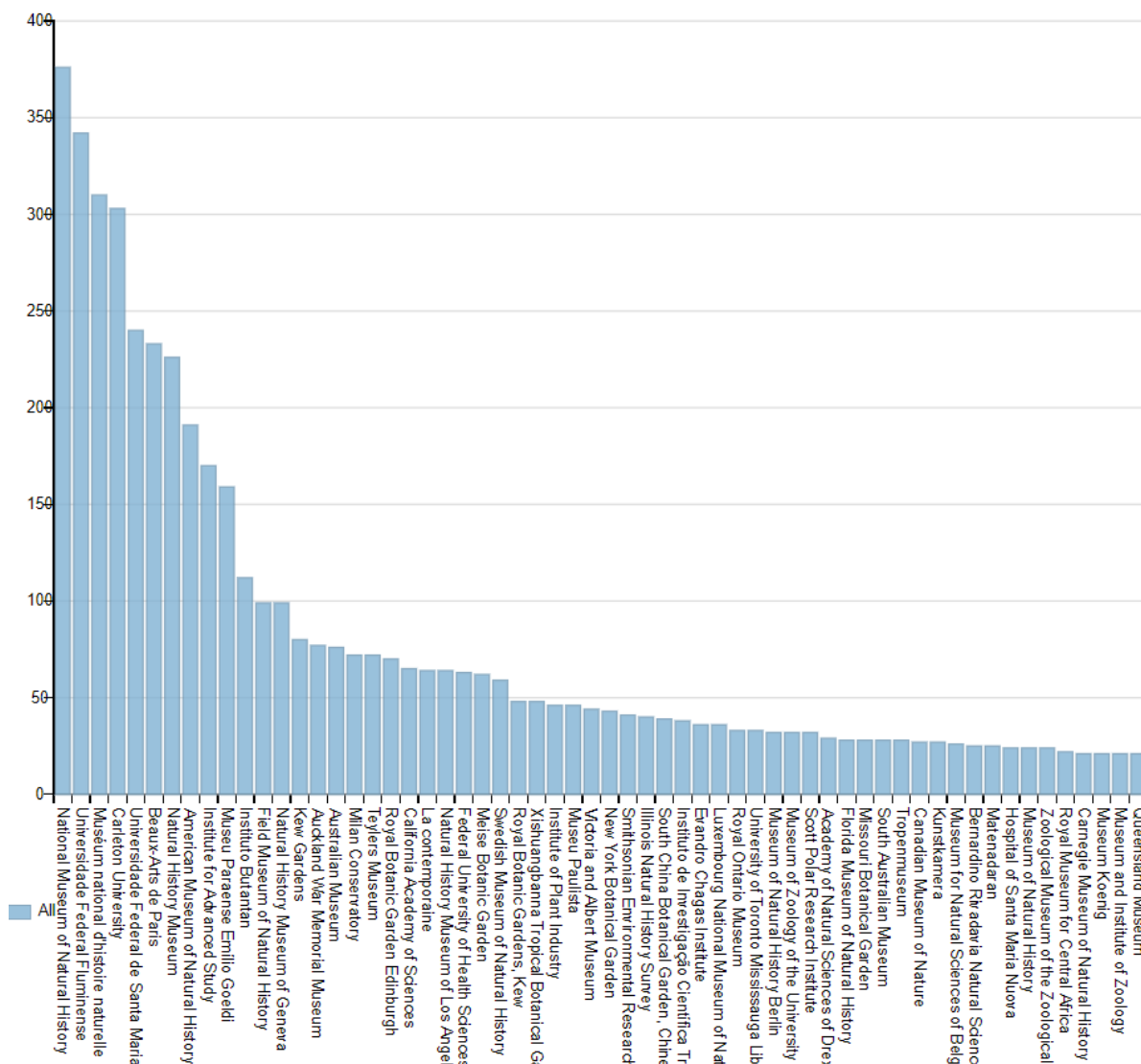
An [example of a SPARQL query that tallies people employed by a natural history museum, museum, botanical garden or archive grouped by institute](#) (top 60 in figure, number of employees on Y-axis.)

personLabel	employerLabel	start	end
Marc E.H. Jones	Natural History Museum	5 February 2018	4 February 2019
Norbert Holstein	University of Bonn	1 October 2014	30 August 2018
Xiaoya Ma	Natural History Museum	1 February 2011	31 May 2018
Tom S. White	University of Cambridge	1 February 2016	1 January 2018

Jennifer Harris	Natural History Museum	1 January 2017	1 January 2018
Peter M. Grindrod	Birkbeck, University of London	1 July 2013	28 July 2017
Marc E.H. Jones	University of Adelaide	30 June 2013	29 June 2017
Marc E.H. Jones	University College London	1 April 2007	1 February 2017

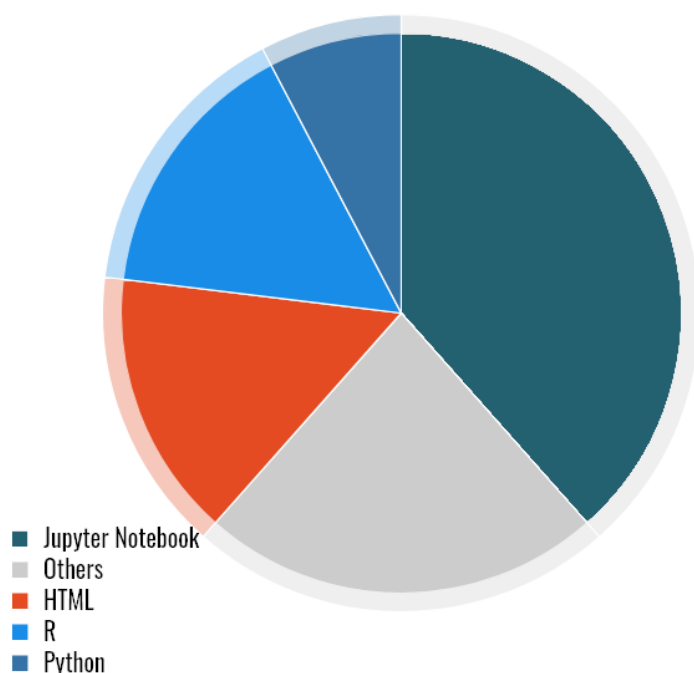An example of a SPARQL query that returns employment history of Natural History Museum staff
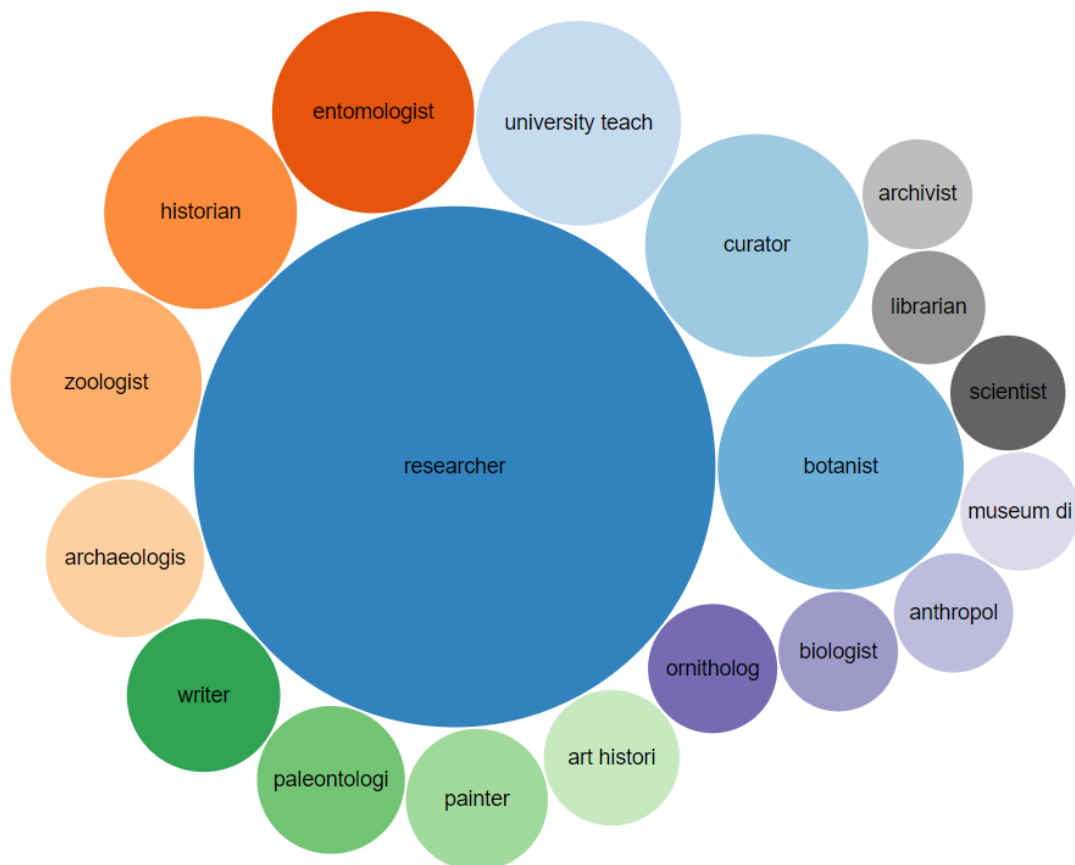
Common professions in Wikidata for people who either were or are employed by natural history museums, botanic gardens, museums, and archives, bigger circles are more common occupations: https://w.wiki/yX$

3.2.4 GitHub

GitHub is increasingly used by scientists to maintain an online portfolio of ongoing and past projects. The platform allows users to query public profiles and their hosted repositories for valuable information such as the used tools and programming languages. However, its coverage remains too small for a full institutional assessment. Furthermore, many users publish on different profiles than their own (such as specific profiles for a project), or for legitimate reasons might "branch" repositories to their own profile that they have no or very limited authorship of. These factors add significant complexity to interpreting the information originating from GitHub, even though it has an open and well documented API[7].

This is an overview of the most used programming languages on the Meise Botanic Garden GitHub profile:

- ■ Jupyter Notebook
- ■ Others
- ■ HTML
- ■ R
- ■ Python

http://ionicabizau.github.io/github-profile-languages/?user=AgentschapPlantentuinMeise

3.2.5 ORCID

Like LinkedIn, and contrary to Wikidata, ORCID offers total control of the content of a profile to its owner. With permission of the owner it can be automatically added to by other services, particularly for publications.

The available information on ORCID is mainly focused on disambiguation of people, and includes employment and educational information, as well as grants and publications. It also holds links to other online profiles including LinkedIn and GitHub.

The usage of ORCID ID as a personal identifier by journals and funding bodies offers an incentive to researchers to maintain their profile. This incentive does not extend to staff that are less involved in this part of the academic process, such as database managers or collections technical staff. ORCID

---

[7] https://docs.github.com/en/rest

explicitly discourages matching staff IDs based on a search by name[8], but instead suggests researchers provide their ORCID ID via OAuthentification. When a list of IDs is built, retrieving public information is trivial via the Public API, however to gain access to non-public fields Trusted Access[9] is necessary which requires permission from the profile owners. This last process could be integrated in the authentication process when researchers provide their ID.

3.3 Linking job titles to skills: enrichment

When the job title of a person is known it is possible to infer a certain set of skills and qualifications for that position. Public datasets exist that help make these assumptions. The Open Skills Project[10] publishes an API that facilitates the normalisation of job titles with common skills and frequently used tools. The underlying information is based on an expansion of the US Department of Labor's O*NET data resources, which are also publicly available[11]. Applying this sort of enrichment to competency analysis opens up two underlying issues. First, if we infer very specific skills and toolsets for a position, the likelihood of error is high. If you take the opposite approach, and only infer very general skills or qualifications, the value of the enrichment is rather low.

3.3.1 Mining publications

Through platforms such as ORCID it is trivial to build a large corpus of publications associated with a person or institute. In principle these could be textmined for certain keywords that indicate fields of activity, or research capabilities. However, retrieving the full text of these publications might form an obstacle, and a significant time investment will be necessary to map full text to a dictionary of interested terms such as competencies and tool sets. A further limitation are the difficulties to ascertain each author's contributions to a publication, especially when many different institutes and authors cooperate on a single work.

Instead of mining full texts for this information, a similar exercise is also possible on publication abstracts. These abstracts can be retrieved using several APIs such as the one Crossref makes available[12]. The viability of such an endeavor is of course dependent on the population of the abstract field in Crossref.

3.3.2 Paid profile enrichment

Several companies offer so-called '*profile enrichment*' as a service. When provided with one or more identifying pieces of information, such as names, email addresses, company name, phone numbers, or social media URLs these companies return other information they store in local databases for a fee. One such service is offered by People Data Labs[13] in the form of the Person Enrichment API[14]. The applications of this service include candidate selection and recruitment, directed marketing, fraud and identity detection, and workforce planning.

---

[8] https://info.orcid.org/faq/how-do-i-find-orcid-record-holders-at-my-institution/
[9] https://info.orcid.org/documentation/features/member-api/access-trusted-data/)
[10] http://dataatwork.org/data/
[11] https://www.onetonline.org/
[12] https://www.crossref.org/education/retrieve-metadata/rest-api/
[13] https://www.peopledatalabs.com/
[14] https://www.peopledatalabs.com/enrichment-api

This service could be employed to enrich publicly available information about staff members across the sphere of collections holding institutes with information about past employment, education and experience (competencies, skills, tools). People Data Labs also offers APIs that can transpose job titles into competencies, and retrieve information about companies such as active industries and associated staff profiles. A likelihood score is computed for every match, and can be used to control the trade off between getting a high response rate or getting fewer but higher precision responses.

People Data Labs offers 1000 requests for free, after this, the Enrichment API costs $0.25 per match. Volume discounts, custom or complete dataset purchases are possible.

The fields provided differ by company, but often include current and past employment, contact information including social media profiles, education and certifications and specific experience. Some companies specialize in parsing out this information from internal documents, and then enriching it through analysis pipelines which can include local datastores. Textkernel[15] offers several (paid) APIs[16] that can extract tools and skills from human resources documentation such as performance reviews, curriculum vitae and vacancies. Furthermore it offers other workforce planning services aimed to pivot people from old to new projects based on information that is locally stored in the institution, but might be difficult to extract.

3.4 Searching for individual competency data - conclusions

No one service was found that can provide all the information required for a dashboard or tool to search individual competencies. It is unlikely that one public entity could be populated and hold all this information within the confines of each platform's limitations. Therefore collating information from different sources seems to be the most productive approach. Central to such an approach stands a unique and persistent identification of people such as the one provided by ORCID. Many staff members of DiSSCo partners will already have an ORCID record due to the requirements of funding agencies and journals. Further provision of services incentivizing ORCID uptake, and automatic linking between publication repositories, can increase the coverage of ORCID in the future.

We conclude that it is only possible to gather information about skills and competencies from a variety of different sources. These sources include data provided by the person, such as LinkedIn and ORCID; provided from elsewhere, such as the People Data Labs datasets; or a combination of both user provided and public sourced, such as ResearchGate. The information from these sources can be further enriched on the personal level by using institutional human resource information services, such as Text Kernel, or on the job-title level using open datasets, such as the one offered by the Open Skills Project.

The cost and complexity of such approaches mean that they will not be pursued further within this Task 3.1, however a combination of increased transparency of internal and external information regarding capabilities and the additional capability to disambiguate persons via an external persistent identifier has the potential to facilitate cooperation and knowledge sharing between DiSSCo partners in future if desired.

---

[15] https://www.textkernel.com/
[16] https://www.textkernel.com/solutions/

DiSSCO
PREPARE

04 Investigation of digital maturity tools

As identified in our Milestone 3.1 report, there are fewer digital maturity frameworks available than competency frameworks. Two, however, were identified as most relevant to our sector and to a potential DiSSCo tool – the Global Biodiversity Information Facility (GBIF) Capacity self-assessment, and the UK Arts Council's Digital Culture Compass Tracker tool. These are examined in more detail in this chapter.

4.1 The GBIF Capacity self-assessment

As described in the source page, the GBIF Capacity self-assessment[17] is a tool to help countries with planning at any stage in the development of their national biodiversity information facility. The self-assessment is related to the GBIF guidelines to establish an effective participant node[18]. It helps to assess capacities that underpin sustainable progress in four key areas:

- **Coordinating a national community of institutions**, projects and initiatives relating to biodiversity, including making connections to the international GBIF network
- **Mobilizing biodiversity data** so that as many sources as possible are freely and openly available
- **Reusing the available data** to support biodiversity-related science and support decision-making for sustainable development
- Improving **biodiversity data management** and data quality to support users' needs

The tool is implemented as an online survey, organised as a series of 50 steps in four sections, each with:
- One screen for context
- Five or six questions, normally of type: yes/no
- Two sub-questions per question, depending on the answer to the main questions
- A final question allowing for unscored text commentary

Answers and scores are given in three levels: no, partially, and totally, with a final score aggregating all of these into one number.

A persistent link is created for each survey, which can be revisited for review/update of the capacity of the organization, and a pdf output can be generated. GBIF requires the exercise to be made by Nodes or institutions that apply for grants. In this way, institutions are incentivised to understand their capacity gaps with the self assessment.

---

[17] https://www.gbif.org/tool/6Y2SqK8XokHUqIFUn6TLxX/online-capacity-self-assessment-tool-for-national-biodiversity-information-facilities (accessed at 14.13 27/04/21), or a previous (deprecated) PDF version is available at https://www.gbif.org/document/82277/deprecated-capacity-self-assessment-guidelines-for-national-biodiversity-information-facilities
[18] https://doi.org/10.15468/doc-z79c-sa53

Example of survey content

The following provides an example of one of the survey sections - **Implementing a national biodiversity data mobilization strategy:**
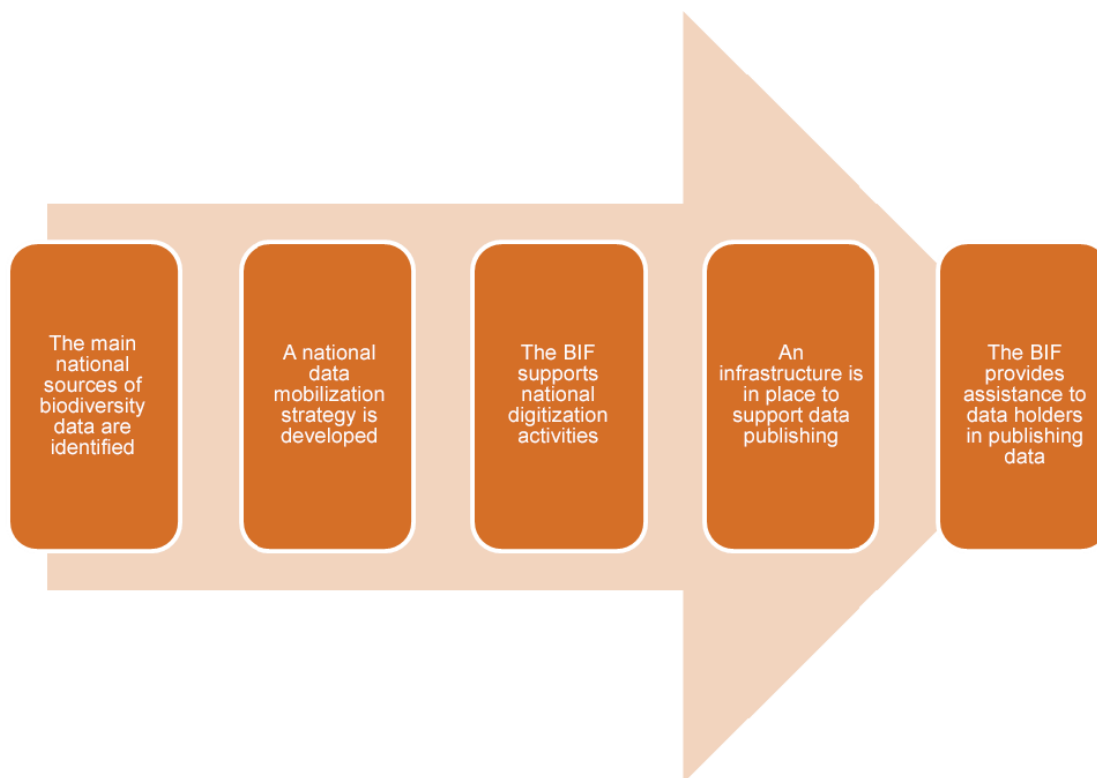
1. Have the main national sources of biodiversity data been identified?
> yes
> no (score 0)

> 1. a. Have the national sources of biodiversity data been documented (e.g. in a metadata catalogue)?
>> 1- Data sources not yet documented (score 1)
>> 2- Basic inventory of data sources is maintained (score 2)
>> 3- National metadata catalogue in place documenting main data sources (score 3)

> 1. b. Has the state of national data holdings been assessed (e.g. digital/non-digital format, approximate size and scope of collections, use of standards)?
>> 1- Data holdings not yet assessed
>> 2- Basic assessment of some holdings
>> 3- Complete and up-to-date assessment of data holdings

2. Has a strategy been developed to assist the biodiversity information facility in mobilizing national biodiversity data sources in a systematic way?
> Yes
> No

2. a. Have data mobilization priorities for the biodiversity information facility been agreed with stakeholders?
    1- Data mobilization priorities not yet identified
    2- Priorities identified but not fully agreed
    3- Priorities identified and agreed

2. b. Does the strategy assist data holders in making a case for investment in the mobilization of their data (e.g. through addressing known gaps or targeting specific use cases)?
    1- Strategy not yet in use
    2- Strategy in use by at least once data holder
    3- Strategy in use by several data holders

3. Does the node (and/or other members of the biodiversity information facility) provide support to national digitization activities?
    Yes
    No

3. a. Are mechanisms in place for national biodiversity data holding institutions to share experience and expertise relating to digitization?
    1- No such mechanisms in place
    2- A few exchange mechanisms
    3- Regular opportunities for exchange

3. b. Does the node promote incentives for data publishing (e.g. through data management policies attached to public research grants, data paper publishing)?
    1- Data publishing not yet actively promoted
    2- Basic incentives promoted
    3- Wide range of incentives promoted

4. Is an informatics infrastructure available to support the publishing of biodiversity data from the country's institutions to the Internet?
    Yes
    No

4. a. Are data hosting facilities available to national holders of biodiversity data?
    1- Options for data hosting not yet available
    2- Data hosting is available for some data holding institutions
    3- Data hosting is offered to all data holders

4. b. Are the data publishing tools provided by the nodal institution and the community kept up to date with the latest tools and software releases?
    1- Data publishing tools not regularly updated
    2- Data publishing tools hosted by the nodal institution are regularly updated
    3- All data publishing tools hosted by the node and community are regularly updated

5. Does the node (and/or other members of the biodiversity information facility) provide assistance to data holders in publishing their data?
    Yes
    No

5. a. Does the node (and/or other members of the biodiversity information facility) organize training for data holders on the topic of data publishing?

  1- Data publishing training not yet organized

  2- At least one data publishing workshop held

  3- Data publishing workshops organized frequently

5. b. Does the node provide a helpdesk service to assist data holders in publishing their data to the Internet?

  1- Helpdesk service not yet in place

  2- Technical assistance provided to some data holders

  3- Efficient helpdesk service in operation

Additional comments or notes to this section (optional - does not affect score) - text box

How might the GBIF approach be relevant for DiSSCo?

The GBIF Capacity self-assessment is focused largely at the national level, as is to be expected given its purpose in relation to national facilities. Many of these areas may not be within the control of a single institution, so we expect the content to vary significantly from what may be required for a DiSSCo digital maturity tool, however there are several relevant features here. The GBIF tool:

- Provides the means to do a standard assessment
- Facilitates the identification of capacities, gaps and needs
- Is autonomous and can help in self planning
- Can combine several aspects of capacity, from governance to technical.

These are all aspects that we will need to replicate in a DiSSCo tool.

The UK Arts Council Digital Culture Compass Tracker

The Digital Culture Compass Tracker tool[19] is for the culture sector, and is not specific to natural history or the natural sciences. It is intended for arts and heritage organisations to assess their current use of digital, and to set 12 month targets, chosen and prioritised to suit the organisation or team. Free registration is required for use.

---

[19] https://digitalculturecompass.org.uk/using-the-tracker/ (accessed 14:17 27/04/21)

The sections of the tool are as shown above. The tracker starts with a customisation section where the user can choose which sections are applicable to their organisation or circumstances. This is a useful tailoring for different organisations that could be relevant for a DiSSCo tool. Users can also toggle a section to 'not applicable' when they are in it, for example if they decide when they see further detail of the questions that something is not relevant to them. This toggle can also be used at question level within a section, which is very useful for more granular tailoring. Sections appear to default to 'not applicable' so appear greyed out, but still allow scores to be entered - as a user this can be confusing and should be avoided.

There are buttons to prompt saving in every section, but autosave also appears to be enabled since the tool can only be completed within a registered account. The tool autosaves for example when 'view assessment' is requested.

Organisation profile includes country and region information; numbers of staff and volunteers; some funding information; and selection of organisation type e.g. museum, arts organisation. After this organisational information, there are a range of sections known as 'activity areas'[20] - each activity area has subcategories with questions for scoring, for example the first element of the Strategy and Governance activity area is Strategy Development and Planning.

Each question includes a statement for each of five standardised levels: **Initial; Managed; Integrated; Optimising; and Transforming**. These statements are tailored to the topic, but are also somewhat standardised in phrasing, for example referring to use of data for planning, or using terms such as 'integrated', at the same level in any topic where that might be relevant. This enables some consistency across levels and questions, but can also mean that the statements feel less relevant or explanatory to some topics, or feel repetitive as a user.

The tool recognises that the levels are not necessarily linear e.g. it's possible for a user to score their organisation more highly against the 'transforming' than the 'managing' level for a particular element. An example in the context of DiSSCo members might be that they are participating in transformational data activity via DiSSCo, but still have some areas for progress in how they are using data internally. Similarly it is possible to target some levels higher or lower than others as you find most relevant for your institution or team.

Scores are self-recorded against each statement for both a current and 12-month target state on a scale of:
0 - not achieved
1- partially achieved
2- fully achieved

An example of one question with the five statements and scoring is show below – this is for the topic 'Strategy Development and Planning':

---

[20] Full details of what is included in the activity areas and subcategories can be found here:
https://digitalculturecompass.org.uk/activity-areas

Notes can be added to each subcategory section, divided into notes on the current position and on the target position.

At any point, users can view their assessment. They can choose to change the marking from draft to final, and if sufficient questions have been scored will be able to see some visualisation and reporting of their position including 'capabilities' such as skills and accessibility that are tagged across different questions in different sections. They can download their assessment including charts as a PDF, or download a CSV file with details of questions and scoring[21].

How might this approach be relevant for DiSSCo?

The Digital Culture Compass tracker is a self-assessment tool, with terms and conditions that allow high level use of data for reporting by the Arts Council - this may well be a suitable model for a DiSSCo tool. It does not allow organisations to compare themselves with one another nor does the tool link out to sources of help and information to support improvement.

---

[21] Examples can be seen here
https://drive.google.com/file/d/1uDgMhDzX7ZuxH-4GLeZXEX_kLjuEIUkx/view?usp=sharing and here
https://docs.google.com/spreadsheets/d/1t6dZtVE3jSk7vRKbbPKHbuJ-yWLi8j6WRb3ngW_Uc_I/edit?usp=sharing

The tool is customisable to different organisations at a section and question level, which may well be a useful feature for a DiSSCo tool that needs to support different teams and organisations with different sizes and priorities. It is likely that we would need a DiSSCo tool to be more tailored to the specific needs of DiSSCo - this tool has relatively little about digitisation for instance, while including public programme and other elements that may be less relevant in the DiSSCo context.A combination of some of the aspects and categories of the GBIF tool with some of these would be a good place to start.

This tool can be completed by a team or organisation, and includes only actions that are within their own power to deliver so it makes less reference to the national context than the GBIF tool. Aspects of this may be useful, though less so at Node level.

The scoring in this tool is relatively simplistic (3 scores) with complexity added by the 5 levels and the number of subcategories - as a user this does not always feel ideal, it is both time-consuming and not always sufficiently nuanced. The ability to add notes partly mitigates this but is less useful for reporting.

Themes such as skills are included in questions across all relevant areas. This is helpful to reflect the different skills in different areas, but makes it harder to get a sense of overall organisational focus on skills, competencies or training.

05   Next steps and conclusion

This report examines further competency and related frameworks; develops work to examine whether and how itis possible to search for individuals with particular competencies, skills and experiences; and looks at two tools that can inform the development of a DiSSCo Digital Maturity self-assessment tool.

In relation to searching for individuals, this paper concludes that no single service can currently provide all the information needed to make this possible, given policy and technical restrictions of data use and user choices about what data they provide and make publicly accessible. As set out in chapter 3 above, collating information from a range of sources linked to a persistent identifier such as ORCID may be feasible for future DiSSCo implementation if desired, and may become more straightforward over time, however the cost and complexity of such approaches mean that they will not be pursued further within this Task 3.1.

Insights from the digital maturity tools examined have been fed into Milestone 3.2 of this task, the Digital Maturity Tool Design Blueprint. That document sets out the next steps to develop that tool, including further consideration of e.g. competency vocabularies and takes into account the examples in this document and in Milestone 3.1.