

Identification of provisional Centres of Excellence for digitisation of European natural science collections

Katherine Dixey[‡], Matt Woodburn[‡], Helen Hardy[‡], Laurence Livermore[‡], Vincent Stuart Smith[‡]

[‡] The Natural History Museum, London, United Kingdom

Corresponding author: Matt Woodburn (m.woodburn@nhm.ac.uk)

Reviewable v1

Received: 19 Aug 2020 | Published: 20 Aug 2020

Citation: Dixey K, Woodburn M, Hardy H, Livermore L, Smith VS (2020) Identification of provisional Centres of Excellence for digitisation of European natural science collections. Research Ideas and Outcomes 6: e57750. <https://doi.org/10.3897/rio.6.e57750>

Abstract

Digitisation of natural science collections is fundamental to the vision for the Distributed System of Scientific Collections (DiSSCo), and given the low proportion of collections digitally accessible, it is proposed that 'Centres of Excellence' be developed to accelerate the creation of digital copies of original specimens. Within the ICEDIG project, a team of scientists from across the consortium explored the concept of Centres of Excellence and have constructed a toolset to help identify these centres to support the development of DiSSCo. This report documents this process and describes the toolset.

Keywords

shared facilities, best practice, digitisation, infrastructure, natural science collections, services

1. Introduction

For the purposes of this paper we have adopted the current Wikipedia definition of a Centre of Excellence (CoE) which is "*a team, a shared facility or an entity that provides*

leadership, best practices, research, support and/or training for a focus area. Due to its broad usage and vague legal precedent, a 'Centre of Excellence' in one context may have completely different characteristics from another." Wikipedia contributors (2019).

Within the context of [DiSSCo](#), this generic definition was refined by ICEDIG partners to describe a team with the appropriate leadership and technology, able to transform a natural science collection into digital surrogates (sometimes called a digital twin) of the original specimens. This might be a photograph, a 3D model, or a media asset (e.g. a video) but will always include structured information (metadata) that remains connected to the digital surrogate however it is used and stored. A DiSSCo 'Centre of Excellence' not only facilitates this process of digitisation but may also manage the lifecycle of these digital objects such as helping to preserve content, protecting the original specimen (especially before and during the digitisation process), improving digital access, and enhancing the return on investment through actions or services on top of the digital surrogate and metadata. A DiSSCo Centre of Excellence may also perform a crucial support role, aiding the digital transformation of DiSSCo partner institutions through the provision of training and best practice, to expand the capacity and capabilities of DiSSCo members.

Previous work by Cocks et al. (2020) surveyed the collections-holding institutes in ICEDIG and offers critical insight into a more detailed list of services that might be offered by a digitisation Centre of Excellence. In addition, a wider survey was reported on by Hardisty et al. (2020) on culture, skills and capacity building. This provides a more detailed analysis of the gaps in skills and/or human resources for digitisation. The results and designs of these surveys have helped to define an initial list of services and capabilities relevant to defining a DiSSCo Centre of Excellence.

This report synthesizes this work to specify a framework for defining models for Centres of Excellence and the organisational levels which are most appropriate to support them. Recognising that a 'one size fits all' approach does not match the strategic, political or financial realities of DiSSCo, we do not nominate specific institutions or facilities as Centres of Excellence. Rather, we provide a framework by which an entity might be assessed in order to qualify as a Centre of Excellence, within the varied context of a specific activity or programme. We also identify clusters of related tasks which might logically be delivered together at different organisational levels, in the creation of these centres.

1.1 Project Context

This project report was written as a formal Milestone (MS45) as part of Task 7.2 of the [ICEDIG Project](#). It was previously made available to project partners and submitted to the European Commission as a report on 29 June 2019. While the differences between these versions are minor the authors consider this the definitive version of the report.

2. Workshops

2.1 ICEDIG All Hands Meeting, Helsinki, 4th June 2019

A workshop was held during the 3rd All-Hands meeting, including representatives from all participating ICEDIG institutes (Suppl. material 1). The goal was to consider what makes a digitisation 'Centre of Excellence' by reviewing three potential hub models (Continental, National, and Institutional) and outlining recommendations for which may be the most appropriate for supporting DiSSCo. The key discussion points were:

- Definition of a Centre of Excellence and the proposed attributes it should offer.
- A review of possible existing examples.
- Target audience for the report: DiSSCo National Task Forces (NTFs), DiSSCo governance & steering group.
- Strengths and weaknesses of proposed models.

It was concluded that digitisation services should be mapped to different hub models, with a non-prescriptive design approach to identifying Centres of Excellence. Subsequent actions included the integration of outputs relating to digitisation capacities from across the ICEDIG project which are summarised by Hardisty et al. (2020). The need to further identify the audience for the milestone, and services which would benefit from centralisation, were also established as open issues.

2.2 Milestone 45 Teleconference, 24th June 2019

This second workshop aimed to review the draft models for DiSSCo Centres of Excellence (described below) which were developed following the ICEDIG All-Hands Meeting. These were presented to meeting participants for feedback and improvement. Points for discussion included:

- Reviewing the list of proposed digitisation services.
- Reviewing the milestone and report structure.
- Use of a more granular service matrix to assess thematic and context-specific models.
- The potential role of commercial organisations within Centres of Excellence.
- Consideration of time-frames: what is expected of a Centre of Excellence in 2, 5 and 10 years and would a Centre operate over a fixed duration (i.e. close after the digitisation of a collection is complete)?

The conclusions and subsequent collaborative review of the draft report have been integrated into a matrix highlighting the appropriateness of delivering a service at different organisational levels.

3. Models for DiSSCo Centres of Excellence

3.1 Services

Drawing on previous work by Cocks et al. (2020), outputs from other ICEDIG work packages (e.g. capacity building in Hardisty et al. 2020) and institutional/community experience, a high-level list of digitisation-related services was identified as potential offerings for a Centre of Excellence. These are summarised in Table 1 with descriptions in Suppl. material 4.

Digitisation	Programme	Infrastructure	Data
<ul style="list-style-type: none"> • High-throughput imaging services • Boutique digitisation • Specimen logistics management • Digitisation cost models • Transcription and translation services • Workflow design • Pre-accession digitisation • Pre-digitisation curation 	<ul style="list-style-type: none"> • Training • Case studies • Funding advice, support and coordination • Building human networks • Communications and advocacy • New workflows and techniques 	<ul style="list-style-type: none"> • Data preservation and storage solutions • Data policies and standards • Data preservation and storage brokerage • Collections management systems • Project and programme skills and tools • Holding and lending specialist equipment 	<ul style="list-style-type: none"> • Quality Assurance • Audience/user insights • Tracking benefits and impact • Data access and discovery platforms (public and/or research) • Data enhancement services • Dealing with aggregators (e.g. GBIF nodes)

3.2 Assessing services against organisational levels

Based on existing digitisation infrastructures globally and within the European scope of DiSSCo, a number of organisational levels were identified at which a Centre of Excellence might operate. Although there is a certain amount of overlap between the different levels, these have been broadly defined, for the purposes of this report, as:

- **Institutional:** a single institution (e.g. Naturalis Biodiversity Centre, NHM London)
- **Regional:** small, geographically linked infrastructures that may have ongoing collaborations, shared programmes and/or services
- **National:** country-level infrastructures (e.g. [DCOLL](#) - Germany, [e-ReColNat](#) - France)

- **Pan-European:** international consortia of any size (2 or more countries) thus encompassing smaller collaborations or larger infrastructures

These four levels were assessed against each of the services outlined in the previous section, and assigned a fitness score from 0 ('inapplicable') to 3 ('high'), with comments on the rationale behind the score. The initial analysis is provided in Suppl. material 2.

This method can be used to identify clusters of services that are a good fit at a particular organisational level. This provides an indication of the different types of Centre of Excellence that may logically be formed and implemented at different organisational levels within the DiSSCo framework.

3.3 Contextual assessments and thematic approaches

While this methodology provides a relatively coarse framework for defining Centres of Excellence, workshop discussions also raised some more nuanced contexts which might influence the fit between services and operational levels. A common case was that of thematic Centres of Excellence, where services focus on specific collections constrained by characteristics such as object type, taxonomy and geographic regions, and the related digitisation workflows and domain expertise. Such specialisms can potentially have influence on factors like funding models, legislative and legal requirements, availability of facilities and logistics that differ from the more generic model. There are also regional contexts to be considered, where the fit between services and organisational levels may be influenced by patterns of local and national funding, institutional expertise and regional differences in collections management practices. This having been said, Centres of Excellence should follow a principle of harmonising differences in practices across thematic, geographic and community boundaries where possible and beneficial.

The variables involved in applying the additional dimensions are too numerous to address within the scope of this work. However, a framework to evaluate each factor involved in matching services to organisational levels in greater detail could be used to structure these more contextual assessments. A basic example of how this might be approached is shown in Suppl. material 3.

3.4 Role of Commercial Organisations

The WP7 group discussed at length the critical role commercial organisations might play within the context of DiSSCo Centres of Excellence. A number of commercial organisations (including ICEDIG partners) have been a key driver in the development of technologies critical to the delivery of high throughput specimen digitisation and in some cases these organisations have become major providers of collections digitisation services. These organisations are likely to continue to play a critical role within DiSSCo, supporting both the innovation of process and the provision of capacity necessary to transform access to natural science collections. Given this, it begs the question as to whether commercial organisations could become DiSSCo Centres of Excellence. The ICEDIG WP7 group concluded that while this was not impossible, any proposal for a commercial entity to

become a DiSSCo Centre of Excellence would need to robustly address four key questions that potentially conflict with the principles of DiSSCo:

1) Breadth of service provision

Commercial organisations often offer a broad portfolio of digital services, but at present, none offer the breadth of provision currently envisaged by a DiSSCo Centre of Excellence. In some cases, this may require investment in processes or service provision where there is no medium or long term financial profitability, making it difficult for a commercial entity to support the activity.

2) Conflicts of interest

Some DiSSCo services (e.g. capacity enhancement) are likely to be in commercial conflict with other aspects of the same organisations work (e.g. digitisation), creating a conflict of interest that would make it impossible for a commercial entity to become a full DiSSCo Centre of Excellence.

3) IP management

DiSSCo, like most EC funded activities, is founded on the principle that investment in intellectual property (e.g. new technologies or processes) is openly available to all. It would be hard to see how this IP protection could be managed by a commercial entity in the context of a Centre of Excellence, as it may give them an advantage not open to the rest of the DiSSCo community. Any exception to this needs to be carefully agreed by all relevant stakeholders prior to the appointment of a commercial organisation as a Centre of Excellence.

4) Sustainability

We envisage some services (especially those relating to data management) are likely to be required in perpetuity. Commercial provision of these services requires careful management to ensure that they are both sustainable and remain commercially competitive over an exceptionally long timeframe. Commercial delivery of these services need to include contractual provision to fully hand over these activities, while mitigating the risks of vendor lock-in associated with the technologies used in delivery of the services.

Given these challenges, at this stage in the development of DiSSCo, the contributors to this report are of the view that it is highly unlikely that a commercial entity could provide the necessary assurances covering all these issues, to qualify as a Centre of Excellence. Despite this we expect commercial organisations to maintain their critical role in the provision of specialist DiSSCo services, through commercial agreements with DiSSCo stakeholders including possible Centres of Excellence.

3.5 Proposed models

The heatmap assessment (Fig. 1) using the service-level matrix reveals clusters of related services with greater suitability at different organisational levels. A summary can be seen in Table 2.

Table 2.

Cumulative fit scores of services clusters and levels, and resulting best-fit model.

Service Cluster	Level				Result	
	I	R	N	P	Highest Fit	Lowest Fit
Digitisation	21	18	18	13	Institutional	Pan-European
Programme	13	14	17	18	National, Pan-European	Institutional, Regional
Infrastructure	11	12	17	13	National	Institutional
Data	10	6	12	18	Pan-European	Regional

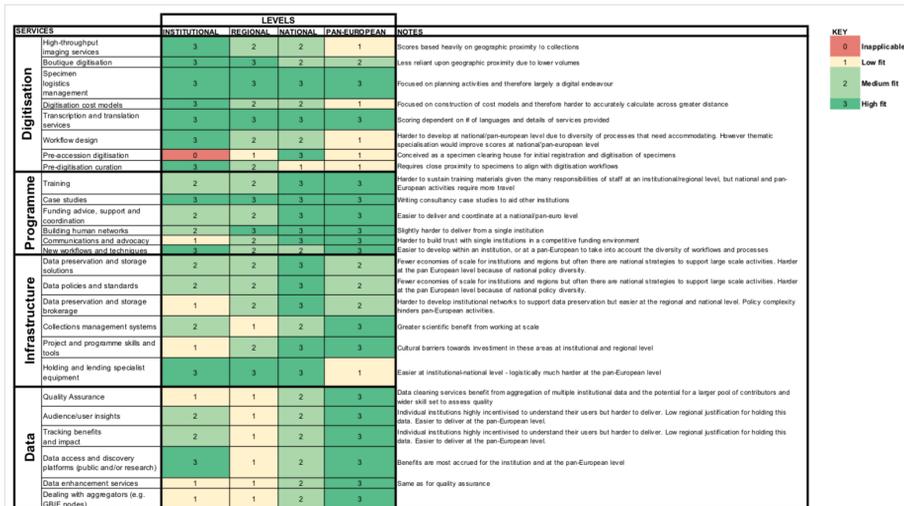


Figure 1. doi

Heat map matrix of Center of Excellence services versus organizational levels.

Operational **digitisation** services (see Table 1) demonstrate increased suitability at the institutional and to some extent regional levels. These services are typically facilitated by geographic proximity of collections and infrastructure. Established cost models and workflows need to accommodate diverse processes and therefore tend to require customisation at the institutional level. Two exceptions which score highly at all levels are *specimen logistics management* and *transcription and translation* services. The latter may benefit both from a distributed range of greater and varied resources, as well as from localised scientific and/or linguistic specialisations. The only service considered

inapplicable at the 'institutional' level was *pre-accession digitisation* - proposed as a 'clearing house' model to digitise specimens before they are received by institutions as by definition it would not be possible and lends itself to a national model.

Programme services encompass training programmes, funding support, networks and communications, as well as development of case studies and new workflows. This cluster was generally biased towards a pan-European or national model. Distribution of resources across a network aids the sustainability of training programmes, whilst advocacy and networking activities are also aided by a distributed model. Funding is facilitated by the existence of collaborations and infrastructures, and a greater range of funding sources can be accessed by linked networks. The development of new workflows and techniques was acknowledged to be easier to develop at the institutional level, but also strengthened by the diversity within pan-European networks. Despite the preference towards more distributed models, all services in this category with the exception of communications and advocacy scored 'medium' or above.

The **infrastructure** service cluster has an identifiable preference towards the national level. National strategies can help facilitate large-scale activities around data storage solutions and standards. Data-related policies also tend to be most effectively applied at a national level, whilst policy diversity at a pan-European level becomes problematic. An exception in this category is holding and lending of specialist equipment, which would experience significant logistical barriers at wider geographical ranges. However, transnational access schemes run by pan-European networks enable sharing of infrastructure and facilities.

The **data** services cluster shows the strongest preference towards a pan-European model. As digital services, these can effectively be constructed around a distributed digital workplace. These services are also likely to benefit from economies of scale of a centralised model, which would also help to drive harmonisation of data standards, processes and platforms across the DiSSCo membership, and integration with core DiSSCo platforms like the European Loans and Visits System (ELViS) being developed by the SYNTHESYS+ Project (Smith et al. 2019).

4. Conclusions

Based on the workshops and framework design process, we have framed a set of high-level principles in identifying DiSSCo Centres of Excellence for DiSSCo:

- Given the breadth of DiSSCo services and levels of operation, there is no single specification for a DiSSCo Centre of Excellence. Rather, a number of models exist that bundle different digitisation services together, and these may be most logically clustered at different organisational levels depending on the precise combination of services.
- Centres of Excellence may provide generic digitisation services, and / or have a particular focus according to themes such as specialist workflows (e.g., herbarium

- sheets, microscope slides, pinned insects), taxonomic groups (e.g. fish, cryptograms, beetles) or geographic and environmental regions (e.g. South America, marine habitats, polar environments).
- A DiSSCo Centre of Excellence could be based in one physical location; physical but distributed (e.g. a regional network); or even virtual. Different services and organisations units lend themselves to different models of distribution, and in all cases, the benefits profile needs to be evaluated on a case by case basis to determine the appropriate level of operation.
 - Different contexts, such as adding a thematic focus, can influence the fit between services and organisational levels. An assessment using a more granular list of factors can be carried out for these on a case by case basis. Thematic specialisation means that what may be appropriate for one centre, may be inappropriate for another.
 - Flexibility is needed for countries to enable them to define national arrangements recognising, for example, regional requirements, patterns of local and national funding, institutional expertise and regional differences in collections management practices. Most often, these differences occur at the national level.
 - Where possible and beneficial, Centres of Excellence should seek to harmonise differences in practices across thematic, geographic and community boundaries.
 - Commercial entities are unlikely to qualify as DiSSCo Centre of Excellence but are expected to take a major role in DiSSCo service provision, by forming relationships with Centres of Excellence or institutional stakeholders. A Centre of Excellence must be able to act as a neutral broker between DiSSCo stakeholders, without the perception of possible conflicts of interest. These centres are likely to offer expertise, support and even training either free or at-cost to users who would not otherwise be able to take these up.

The scoring process suggested service profiles for four potential types of Centre of Excellence for digitisation within DiSSCo, characterised both by complementarity of services and the operational level to which they might be best suited. A suitable next step would be to consider how these service models might align with DiSSCo's distributed model for governance, infrastructure and coordination, and therefore how Centres of Excellence might be realised within the context of DiSSCo.

Funding program

[H2020-EU.1.4.1.1. - Developing new world-class research infrastructures](#)

Grant title

[ICEDIG](#) – “Innovation and consolidation for large scale digitisation of natural heritage”, Grant Agreement No. 777483

Author contributions

Katherine Dixey: Data Curation, Investigation, Writing – Original Draft. **Matt Woodburn:** Conceptualization, Investigation, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing. **Helen Hardy:** Writing – Review & Editing. **Laurence Livermore:** Writing – Review & Editing. **Vincent Smith:** Conceptualization, Writing – Original Draft, Writing – Review & Editing.

Contribution types are drawn from CRediT - [Contributor Roles Taxonomy](#).

References

- Cocks N, Livermore L, Smith V, Woodburn M (2020) Technical capacities of digitisation centres within ICEDIG participating institutions. Research Ideas and Outcomes 6 <https://doi.org/10.3897/rio.6.e55522>
- 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, Willemse L (2020) Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1. Research Ideas and Outcomes 6 <https://doi.org/10.3897/rio.6.e54280>
- Smith V, Gorman K, Addink W, Arvanitidis C, Casino A, Dixey K, Dröge G, Groom Q, Haston E, Hobern D, Knapp S, Koureas D, Livermore L, Seberg O (2019) SYNTHESYS+ Abridged Grant Proposal. Research Ideas and Outcomes 5 <https://doi.org/10.3897/rio.5.e46404>
- Wikipedia contributors (2019) Center of excellence. https://en.wikipedia.org/w/index.php?title=Center_of_excellence&oldid=901202628. Accessed on: 2019-6-21.

Supplementary materials

Suppl. material 1: WP7 workshop for identification of provisional centres of excellence for digitisation [doi](#)

Authors: Matt Woodburn

Data type: PowerPoint Presentation

[Download file](#) (707.92 kb)

Suppl. material 2: WP7 MS45 Centres of Excellence - Service x Levels [doi](#)

Authors: Katherine Dixey, Matt Woodburn

Data type: Excel Spreadsheet

[Download file](#) (81.96 kb)

Suppl. material 3: WP7 MS45 Centres of Excellence - Service Requirements 

Authors: Katherine Dixey

Data type: Excel Spreadsheet

[Download file](#) (12.13 kb)

Suppl. material 4: WP7 MS45 Centres of Excellence - Service Descriptions 

Authors: Katherine Dixey

Data type: Excel spreadsheet

[Download file](#) (10.34 kb)