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
DPP WP4 D4.1 "The Cost Book for DiSSCo"

Author(s)
Salomé LANDEL
Michel GUIRAUD
Ana CASINO

Identifier of the author(s)
Salomé LANDEL 0000-0001-5360-5693
Michel GUIRAUD 0000-0003-3125-8947
Ana CASINO 0000-0002-9869-6573

Affiliation
Muséum national d'Histoire naturelle (Paris)

Contributors
Natural History Museum (London)
Naturalis Biodiversity Centre (Leiden)
Muséum national d'Histoire naturelle (Paris)
Consortium of European Taxonomic Facilities
Senckenberg Gesellschaft für Naturforschung (Frankfurt)
Meise Botanic Garden
Royal Belgian Institute of Natural Sciences (Brussels)
Botanischer Garten und Botanisches Museum (Berlin)
Natuurhistorisch Museum Maastricht
Natuurmuseum Brabant
Royal Netherlands Institute for Sea Research
Università degli Studi di Firenze
Naturhistorisches Museum Vienne
Université Claude Bernard Lyon 1
Univeristy of Tartu National History Museum and Botanical Garden
Museum für Naturkunde Berlin
Teylers Museum
Muséum d'histoire naturelle de Marseille
Université de Rennes 1
Muséum d'histoire naturelle de La Rochelle
Muséum d'histoire naturelle de Nice
Centre de coopération internationale en recherche agronomique pour le développement (France)
Naturkundemuseum Stuttgart
Institut de recherche pour le développement (France)
University of Copenhagen, Natural History Museum of Denmark
Université de Bourgogne

Publisher

Identifier of the publisher

Muséum national d'Histoire naturelle (Paris) grid.410350.3

Resource ID

Publication year
2023

Related identifiers
https://doi.org/10.1093/biosci/biac060
https://doi.org/10.3897/rio.7.e67379
https://doi.org/10.3897/rio.6.e58915
https://doi.org/10.3897/rio.6.e54280
https://doi.org/10.3897/rio.8.e93816
http://doi.org/10.5334/dsj-2020-050
https://doi.org/10.34960/1xr6-hr45
https://doi.org/10.3897/rio.6.e56211
https://doi.org/10.3390/info12110472

Relation type
Another reference document

Is it the first time you submit this outcome?
Yes

Creation date
06/01/2023

Version
1

Citation
Landel, S. ; Casino, A. ; Guiraud, M. (2023) The Cost Book for DiSSCo. DiSSCo Prepare WP4 - D4.1.

Abstract
Research Infrastructure (hereinafter referred as to RI) cost calculation can be multifaceted and complex. DiSSCo is at the same time a central team and a coordinated network responsible for supplying the infrastructure's services. The RI linked costs are spread all over Europe and are connected to thousands of people. As the extended DiSSCo perimeter encompasses a wide range of services - from physical access to digitisation on demand and consulting services - distributed among a great number of partners, cost information is significantly decentralised.

The first "centralised" cost calculation leads to the following results:
• The DiSSCo Central Hub office would need a minimum annual budget of €M 1.4 to be operational. This would not change between the construction and the operation phases.
• The DiSSCo Central Hub IT team would need a budget of €M 2.2 to finalise all IT system under the construction phase (if the construction phase lasts two years, it would cost around €M 1.1 per year) in order to develop the digital services that will facilitate access to NSC data. The annual cost to operate these services would be around €M 1.2 per year.

This budget will evolve according to funding opportunities, the enlargement of the membership and the implementation of a business model open to new sources of income. As it is often the case, research infrastructures grow over time and the more DiSSCo will be known and recognised, the more it will attract users, and the more its budget will increase.

In order to calculate the RI linked costs, which are spread all over Europe, T4.1 developed a cost calculation methodology that has been distributed among all the 170 DiSSCo partner institutions. 27 institutions responded to the exercise. It allows for a first shared understanding on how to calculate DiSSCo related costs. It also provides the first figures on a cost per hour or a cost per service. It opens the door for a pricing under the realm of DiSSCo. Finally, such a methodology also aims at guaranteeing a fair service pricing based on the same principles and variables.

Content keywords
financial

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP4

Project output
Deliverable

Deliverable/milestone number
4.1

Dissemination level
Public

Rights

Resource type
Text

Format
PDF

Funding Programme
H2020-INFRADEV-2019-2

Contact email
salome.landel@mnhn.fr

H2020-INFRADEV-2019-2
Grant Agreement No 871043

The Cost Book for DiSSCo

DiSSCo Prepare WP4 – Deliverable 4.1

WP Leader: Michel Guiraud (MNHN)

Authors: Salomé Landel (MNHN), Michel Guiraud (MNHN), Ana Casino (CETAF)

Contributors: Lisa French, Laurence Livermore, Helen Hardy (NHM),
Eva Alonso, Wouter Addink, Sharif Islam, Dimitris Koureas (Naturalis),
Katharine Worley, François Dusoulier (MNHN),
Marko Lovrić (CETAF),
Claus Weiland, Hanieh Saeedi (Senckenberg Gesellschaft für Naturforschung),
Stefaan Pijls, Frederik Leliaert, Patricia Mergen (Meise Botanic Garden),
Carole Paleco, Serge Scory, Patrick Semal, Gaël Lymer (Royal Belgian Institute of Natural Sciences)

Deliverable references

Deliverable	D4.1: The Cost Book for DiSSCo
Work Package	WP4: DiSSCo business framework
Lead Partner	Muséum national d'Histoire naturelle
Status	Draft
Deliverable type	Report
Dissemination level	Public
Due date	6/01/2022
Submission date	6/01/2022

KEY WORDS

Cost calculation, Construction phase, implementation phase, staff costs, operating costs, capital costs, costing methodology, Central Hub, Cost per hour, Service provision, Pricing

Abstract

The European Research Infrastructure DiSSCo (Distributed System of Scientific Collections) aims to digitally unify all European natural science assets, to ensure that collection data are easily findable, accessible, interoperable and reusable (FAIR). 170 institutions across more than 23 countries are involved in this ambitious objective of transforming a fragmented landscape of collections into an integrated knowledge base, enabling researchers to use and interconnect different collections.

Research Infrastructure (hereinafter referred as to RI) cost calculation can be multifaceted and complex. DiSSCo is at the same time a central team and a coordinated network responsible for supplying the infrastructure's services. The RI linked costs are spread all over Europe and are connected to thousands of people. As the extended DiSSCo perimeter encompasses a wide range of services - from physical access to digitisation on demand and consulting services - distributed among a great number of partners, cost information is significantly decentralised.

The first "centralised" cost calculation leads to the following results:

- The DiSSCo Central Hub office would need a **minimum annual budget of €M 1.4** to be operational. This would not change between the construction and the operation phases.
- The DiSSCo Central Hub IT team would need a **budget of €M 2.2** to finalise all IT system under the construction phase (**if the construction phase lasts two years, it would cost around €M 1.1 per year**) in order to develop the digital services that will facilitate access to NSC data. The annual cost to operate these services would be around **€M 1.2 per year**.

This budget will evolve according to funding opportunities, the enlargement of the membership and the implementation of a business model open to new sources of income. As it is often the case, research infrastructures grow over time and the more DiSSCo will be known and recognised, the more it will attract users, and the more its budget will increase.

In order to calculate the RI linked costs, which are spread all over Europe, T4.1 developed a cost calculation methodology that has been distributed among all the 170 DiSSCo partner institutions. 27 institutions responded to the exercise. It allows for a first shared understanding on how to calculate DiSSCo related costs. It also provides the first figures on a cost per hour or a cost per service. It opens the door for a pricing under the realm of DiSSCo. Finally, such a methodology also aims at guaranteeing a fair service pricing based on the same principles and variables.

Abbreviations

AAI:	Authorisation and Authentication Infrastructure
AE:	Accounting entity
CDD:	Collection Digitisation Dashboard
CETAF:	Consortium of European Taxonomic Facilities, Brussels
DG:	Director General
DiSSCo KB:	DiSSCo Knowledge Base
DiSSCo:	Distributed System of Scientific Collections
DO repo:	Data repository for digital specimens
DOI:	Digital Object Identifier
DPP:	DiSSCo Prepare Project
EC:	European Commission
ELViS:	European Loans and Visits System
ERIC:	European research infrastructure Consortium
ESFRI:	European Strategy Forum for Research Infrastructure
EU:	European Union
GA:	General Assembly
ICEDIG:	Innovation and consolidation for large scale digitisation of natural heritage
IT:	Information Technology
MNHN :	Muséum national d'Histoire Naturelle, Paris
MoU:	Memorandum of Understanding
Naturalis:	Naturalis Biodiversity Centre, Leiden
NHM:	National History Museum, London
NN:	National node
NSC:	Natural Science Collections
ORCID:	Open Researcher and Contributor ID
RECOLNAT:	Réseau national des collections naturalistes
RI:	Research infrastructure
SDR:	Specimen Data Refinery
SGN:	Senckenberg Gesellschaft für Naturforschung, Frankfurt
SLA:	Service Level Agreement
SYNTHESYS+:	Synthesys of Systematic Resources
UCAS:	Unified Curation and Annotation System

INDEX

Introduction

DiSSCo, the Distributed System of Scientific Collections, is a distributed European Research Infrastructure (RI) unifying access to Natural Science Collections (NSCs) and delivering it to scientific communities and beyond. Bringing together 170 institutions across 23 countries and combining earlier investments in data interoperability practices with technological advancements in digitisation, cloud services and semantic linking, DiSSCo aims to make the data from NSCs available as one virtual data cloud, connected with data emerging from new techniques and not already linked to specimens.

The DiSSCo Prepare project (DPP) WP4 is dedicated to the business framework of the DiSSCo Research Infrastructure. The business perimeter of DiSSCo is extensive, gathering many stakeholders in a process of transition from a fragmented landscape to an integrated research infrastructure.

This deliverable – D4.1, DiSSCo Cost Book – aims at listing and quantifying DiSSCo future costs. Here, what we mean by Cost Book is "a list of unit costs for the main cost items of the RI (e.g., cost of the access for one day, unit cost of administrative/scientific/managerial personnel, etc.)."[1]For DiSSCo, it requires to study the centralised and decentralised costs of the RI. It is a tool to forecast future costs and to identify how to calculate the price of services linked to NSCs. The DiSSCo Cost Book excludes a detail analysis of all the costs related to NSCs, it identifies methods of calculating costs in a way that is achievable the project members.

Such a task requires to first delimit the perimeter of the RI and the associated human resources. As DiSSCo services are spread among its 170-member institutions, cost calculation needs a methodology allowing the participation of a wide range of stakeholders, with different levels of financial skills. Indeed, the DiSSCo perimeter crosses some established perimeters: it is a puzzle of different pieces of activities within different institutions, departments, service units, etc. Understanding that means that it is often not possible to directly extract financial information from accounting systems. It requires qualitative work and a structured methodology to identify these costs. With an extensive understanding, if all partners share their system and rules on cost calculation, it allows for a transparent price calculation at European level. The first intuition was that the differentiating variable would then be the price differences among countries based on purchasing power parity.

In that context, this document will first introduce the business framework and the RI perimeter. These two narratives provide the context for understanding the deliverable. Following that, the first chapter resumes previous work on cost calculation and explains the initial finding for DiSSCo Cost Book. The second chapter explains in detail the development of the methodology to assess DiSSCo costs. Finally, the third chapter provides the first results of this study.

[1] ESFRI methodology: https://www.esfri.eu/sites/default/files/StR-ESFRI2_STUDY_RIs_COST_ESTIMATION.pdf

DiSSCo RI Perimeter

Each step taken towards cost calculation is impacted by the definition of the RI perimeter. Under the ICEDIG project, the business question was mainly limited to digitisation programmes, meanwhile at the very beginning of the WP4 the perimeter was very broad and encompassed all activities related to DiSSCo.

When it comes to cost calculation, it is important to identify the people who will be able to provide relevant information[2] on the different activities. Such identification helped WP4 to better define the DiSSCo perimeter. In parallel, there is a methodology provided by ESFRI on cost calculation for research infrastructures. This methodology sets a clear limit on the perimeter: '*costs must always be accounted for according to an incremental approach, which requires to consider the costs occurring for the set-up and implementation of the RI not of those that would occur in any case in absence of the RI (the 'without the RI', or counterfactual, scenario)*'**. It means that the costs that have to be calculated are the ones that are created by the existence of the RI itself.**

With that concept, WP4 understood that DiSSCo members (institutions) have activities – with costs – which exist with or without the RI. These activities are costly but the RI will be a bridge between these activities and users. **The costs to be calculated are therefore the costs of "the bridge": a central hub with a dedicated team which coordinates, manages and offers e-services.**

Still, in order to better understand the costs of the activities carried out by the members, and potentially help them to develop tariffs, WP4 designed a methodology to calculate the costs of Natural Science Collections-related activities. This is first based on an understanding of the activities. They are listed below:

1. **Mass digitisation**: is the implementation of a workflow to digitise a large number of specimens in a limited time. It includes the preparatory phase of digitisation up to the production and publication of the data. Indicative threshold: more than 5 000 digitised specimens in a year (up to tens or hundreds of thousands in some cases). Digitisation does not automatically mean imaging, it can only be databasing.
2. **Small set digitisation**: In contrast to mass digitisation, this digitisation is done on a regular basis and/or on demand, and involves only a smaller number of specimens at a time. Indicative threshold: less than 5 000 digitised specimens in a year. Digitisation does not automatically mean imaging, it can only be databasing.
3. **Consulting services / expertise**: corresponds to any requests that rely on giving expert advice, for example determining taxonomy. There is a lot of collection information which is not available through an externally accessible database: collections are not fully digitized, both quantitatively and qualitatively. Some of this information often relies upon the knowledge and experience of collection managers and curators. Consulting services can be the direct interaction with collection managers and curators.
4. **Training**: courses where the participation of experts from Natural Science institutions is crucial. Also important is the tooling-up of the scientific community in their different areas of expertise. The courses can address the needs of staff in different stages of their careers, ranging from early-

[2] Mentioned as "contact persons" in the document.

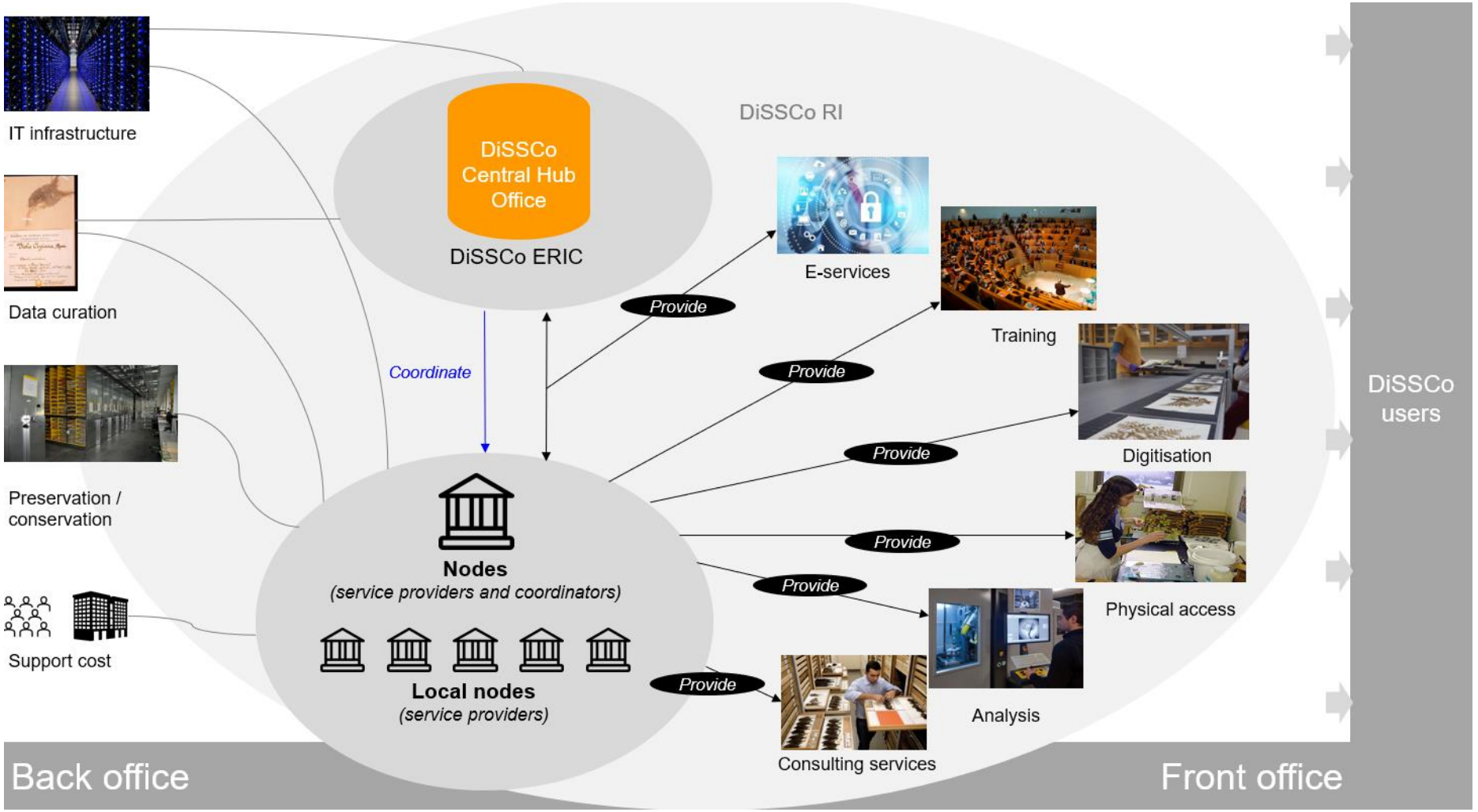
career investigators, young professionals to advanced more experienced senior staff members. They both target RI members and non-members. [3]

5. **Collection analysis services**: produce data that would not be visible to the naked eye. They require a physical intervention on a specimen (sampling, imaging, etc.) and the use of specific analytical facilities (laboratory, microscope, etc.).
6. **Loan of collections**: time accounted for in DiSSCo excludes time spent on loans for exhibition but includes time starting with decision-making committees for these requests, to finding the requested specimens, assessing their condition, sampling fragments if requested, packaging and customs formalities.
7. **Physical access to collections for external researchers**: the time accounted for starts with decision-making committees for these requests, to prepare the specimens for consultation, take care of the material, organization of the visit and deal with the administrative requirement for hosting the visitor.
8. **Asset maintenance**: conservation and data curation. It covers the renovation of collections (relates to work sites, relocation of collections, restoration of a room, etc.), the preventive and curative preservation (day-to-day maintenance of collections to ensure their long-term preservation) and data curation activities (time spent to maintain, manually or computationally, information about natural science collections. It involves updates of digitised information).

The DiSSCo RI perimeter could then be represented with concentric circles (see figure 1). At the centre there is the hub which coordinates the demand and the provision of services. Around this hub, there is the perimeter of the ERIC: it can encompass other bodies than the hub. SLAs might be signed with institutions who would then become service providers on behalf of the ERIC. The final circle is the Research Infrastructure: inside it are the institutions which implement the eight categories of activity listed above. These activities are important for the Research Infrastructure, but would nevertheless be part of the institutions' mission without its existence. The addition of all the concentric circles represents the economic value and, at the same time, the economic impact of the RI.

[3] Definition from Synthesys+ - D2.3 – catalogue and recommendations for the development of a proactive, efficient and evolving DiSSCo training programme.

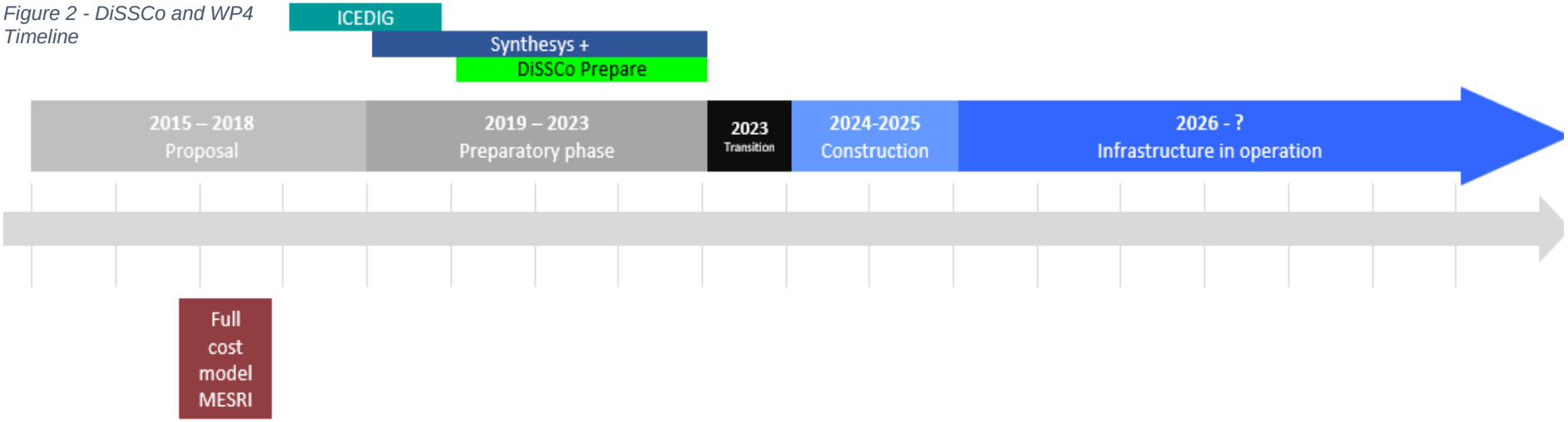
Figure 1 - Illustration of DiSSCo Perimeter



CHAPTER 1: previous experiences inspiring WP4

Prior to DiSSCo Prepare and WP4, other projects took place at European and national level, two of which have influenced the WP4 strategy. The first of these was a survey implemented by the French Ministry of Research (MESRI) in 2016 and 2017. It aimed to calculate the costs of all research infrastructures operating in France. At MNHN level, the national RI RECOLNAT[4] was one of the entities from which the costs were collected. This work had an impact on WP4's vision of cost calculation. Secondly, the European project ICEDIG (Innovation and consolidation for large scale digitisation of natural heritage) had the objective to "set up the necessary technological, socio-cultural, and organisational features to enable the operation of a unified access point to bio and geo-diversity data", which is now designated as DiSSCo. A shared conclusion of both experiences is that **calculating costs always involves a certain degree of uncertainty.** The way the methodology and the concepts used are understood has a considerable impact on the results. Figures are not always fully stable. As Research Infrastructures are often distributed over cross-cutting perimeters[5], the identification of their costs is highly complex. Nevertheless, from an economic viewpoint**, a good understanding of the costs related to the activities implemented at institution-level can improve their efficiency**. As good practice, ongoing recording of the costs can help to develop more cost-effective projects.

Figure 2 - DiSSCo and WP4 Timeline



[4] The National Network of Natural History Collections (RECOLNAT) is a French research infrastructure (RI). It concerns all-natural history collections and their valorisation by research. At the heart of its missions is the production and provision of a corpus of data for the study of current and past geology and biodiversity.

[5] Cross-cutting perimeter: their perimeter of action is distributed among different parcels of service or research units. It means that it is not possible to directly extract financial data from the different units. It is a matter of qualitatively identifying expenditure that relies to DiSSCo from the ones who do not.

1.1 ICEDIG: a first step toward the calculation of Natural Science Collections digitisation costs

Between January 2018 and March 2020, a European Project, ICEDIG, was implemented to pave the way for DiSSCo Prepare and, in the same vein, DiSSCo Research infrastructure. ICEDIG aimed at preparing mass digitisation of European Natural Science Collections and the subsequent access to all related data in a harmonised and integrated manner across Europe. In the framework of DiSSCo Prepare WP4 (Business framework), three main deliverables from ICEDIG feed into the business framework of DiSSCo RI:

- Deliverable 8.2: Cost Book of the digitisation infrastructure of DiSSCo[6]
- Deliverable 4.5: a cost analysis of transcription systems[7]
- Deliverable 8.1: conceptual design blueprint for the DiSSCo digitization infrastructure[8]

1.1.1. Costs study: lack of information on digitisation and transcription systems costs

Deliverable 8.2, Cost Book of the digitisation infrastructure of DiSSCo, is a study of 22 Cost Books from 6 institutions[9] across Europe. Following their experience in the digitisation of Natural Science Collections (NSC), they were asked to break down the associated costs into three categories (capital costs, fixed costs and variable costs). Additional questions were raised notably on staff, throughput and time spent on each specimen.

The results show that there is much more data on some collections (herbarium) than on others (vertebrates). Understanding that, the deliverable also explains that vertebrates and marine invertebrates are much more expensive to digitise than other collections (herbarium and pinned insects). Regarding the category of costs, at the time, mass digitisation was implemented mainly in-house, which includes capital costs. Some institutions chose to outsource the process: a per item cost or a total negotiated price was paid to cover the variable costs of digitisation, recoupment of contractor's capital and fixed costs and provide a profit margin. It is possible to demarcate the two different steps: (i) establishing a digitisation facility is mainly capital costs, (ii) digitising specimens is mainly operating costs.

To calculate costs often means understanding the different steps implemented to accomplish a task. Regarding mass digitisation of NSC, 5 main activities were identified: pre and post digitisation curation; specimen image capture; image processing; data capture; and preserving and publishing data.

---

[6] Hardisty A, Livermore L, Walton S, Woodburn M, Hardy H (2020) Costbook of the digitisation infrastructure of DiSSCo. Research Ideas and Outcomes 6: e58915. https://doi.org/10.3897/rio.6.e58915

[7] Walton S, Livermore L, Dillen M, De Smedt S, Groom Q, Koivunen A, Phillips S (2020) A cost analysis of transcription systems. Research Ideas and Outcomes 6: e56211. https://doi.org/10.3897/rio.6.e56211

[8] Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalga A, Paul DL, 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: e54280. https://doi.org/10.3897/rio.6.e54280

[9] APM (Meise Botanic Garden – Belgium), Luomus (Finnish Museum of Natural History), MNHN (Muséum national d'Histoire naturelle – France), UTARTU (University of Tartu – Estonia), NHMUK (Natural History Museum – UK)

The conclusions of this study are that the workflow to digitise herbarium sheets and pinned insects is much more mature than for other types of collections. Labour, or staff cost, can be seen as both a variable and a fixed cost. This is because staff cost may be impacted by the activities of the digitisation process but, conversely, staff can be considered as a fixed cost because salaried employees are paid with or without the digitisation programme, except in the case of outsourcing. Deliverable 8.2 explains that data is not fully reliable, notably due to "the lack of a common standards, data model and vocabularies have been a significant barrier to making these datasets comparable and interoperable".

Deliverable 4.5, a cost analysis of transcription systems, aims to compare the costs associated with label transcriptions of Natural Science Collections. The basis of this approach is that data from natural science specimens remain largely human readable and only accessible physically: by looking at corresponding handwritten, typed or printed labels or registers. The conclusion of this deliverable was that there is still a lack of a full financial breakdown and full data quality information on these workflows. At the time, the advice was that a fast, efficient, cost-effective means of transcribing label data still needed to be developed. It is one of the conditions required to digitise Natural Science Collections on an industrial scale. The deliverable 4.5 encourages institutions to quantify these costs in order to develop more efficient workflows.

1.1.2. Initial broad estimation of digitisation costs

Deliverable 8.1, conceptual design blueprint for the DiSSCo digitisation infrastructure, is the main and final document of the ICEDIG project. It identifies some of the challenges DiSSCo RI will need to overcome in order to be properly implemented. It recalls the ambition of the project: digitising biological collections, which have been gathered over more than 250 years. In a few figures, Europe's collections represent 1.5 billion specimens and, at the time of the deliverable, about 10% were digitally catalogued. According to the same report, with traditional methods, one person can image and completely digitise the data associated with 50 specimens in one working day, with a basic cost of about €5per specimen. As an estimation, it would thus take 100,000-person years to digitize 1 billion specimens and consequently cost €5 billion. In addition to these costs, the storage of petabytes of image data, analysis of images of millions of specimens, and complementing their metadata with other details requires massive computing power and the mining of data from big repositories. At an institutional level, such a strategy is beyond the capacity of most museums.

1.1.3. Estimation of IT budget requirements

Access to Natural Science Collections can be improved thanks to software development. According to D8.1, 8-10 trained engineers fulltime over 4 years (i.e., 32-40 full time equivalent, FTE) plus one senior technical manager/principal IT architect will suffice in order to bootstrap DiSSCo with the software critical for the construction phase. This staff can represent a minimum cost of €M 4; more (€M 6.4) if experienced, skilled engineers are employed. Such figures can be explained by the fact that competing with the private sector for competent software engineers is expensive. Regarding the operation phase, the costs during the first two years are expected to be similar: no reduction in effort is expected due to ongoing maintenance and support needs, as well as the need to introduce new capabilities.

1.1.4. Alternatives to lower the costs

In that frame of ambitious objectives and restricted capacity, the private sector can be seen as a strategic partnership, notably through procurement. Some companies specialise in digitisation equipment (scanners, cameras, imaging technologies, conveyor machinery and other automation, including associated specialized software). Meanwhile, D8.1 states that purely commercial entities are unlikely to be able to provide enough assurances on several fronts to become a DiSSCo Centre of Excellence (DCE). The only solution would then be to subcontract with them but not to fully associate them in the project.

Another solution is to increase capacities thanks to innovation and the pooling of skills and resources at European level. To reduce costs the number of human operators must be reduced as much as possible. A fully automated line where humans are only required to bring the insect drawers to the digitisation line would be ideal. It means that the community needs to find ways to lower average item costs. According to D8.1, imaging millions of these specimens should be viable at an average target cost of around €0.20 per specimen, and transcription of their essential (MIDS-2 level) data should cost less than €0.30 per specimen when properly supported by technological and automated approaches, and/or done in less expensive parts of the world.

The question that remains therefore, is what is DiSSCo's role within that landscape? Firstly, the RI can homogenise good practices and help its members make their workflows more efficient. At the time of ICEDIG, a second role for DiSSCo was envisaged: to operate mass digitisation systems and services. This option is not contradictory with a RI's mission; however, the limits of this role are linked to the nature of its component parts. This activity would come with a cost: DiSSCo could have a centralised facility for mass digitisation which would require the specimens to be moved from the different Museums in Europe. The location and ownerships of digitisation centre(s) has already been questioned.

One solution to improve digitisation capacities is to develop an accurate understanding of the costs linked to the digitisation process. One basic assumption is that optimal digitisation cost is achieved when the volume and availability of specimens ready for digitisation matches the capacity of the digitisation facility. Another assumption is that different purchasing power parity among countries should be considered in the calculation – in addition to the different currencies. DiSSCo data standards will add value to data and thus have an impact on its business charging model.

1.2 Full cost model from French Ministry of Research: a proof of concept for cost calculation

Between November 2016 and October 2017, the French Ministry of Research (MESRI) asked research infrastructures that are part of its RIs roadmap to evaluate their "full cost" according to a harmonised methodology. One achievement of this exercise is that 100% of the RIs responded and communicated their results to the Ministry. This high percentage is notably due to the active monitoring of the people who led the task at institutional level.

1.2.1. General results

Despite an excellent level of response, the results received were sometimes inconsistent. Not all categories were always fully understood. Another default of the methodology was that it was very time-consuming – which was partly due to the fact that it was new: there was an entry cost. In addition, the structure of the methodology has changed during its implementation period. Some partners had already started working with the initial version. Regarding the indirect costs, 56% of the RIs used a 25%

flat rate. 28% used a 28% flat rate according to a mix of different rates used. 15% concretely calculated the costs. The construction and decommissioning costs were insufficiently calculated. One of the conclusions of the study was that annual dismantling costs, on average, never exceed 5% of annual full operating cost. The degree of reliability of the results received was therefore partial. Some expenditures, such as construction and decommissioning, were not considered. The 25% flat rate leads to an increased valuation. In that sense, the full cost calculated in 2016 has a 10% margin of error. The real cost should be above the global amount calculated through this survey. Another factor of uncertainty is the lack of valuation of the spaces occupied by RIs.

Thanks to this methodology, the result achieved in 2016 is that French RIs and French contributions to international RIs represent a full cost of around €M 1,500. The costs are distributed as follows: staff costs represent 39% of the total, operating costs represent 40% of the total and capital costs represent 21% of the total. Decommissioning costs are marginal: around 0.4% (probably below actual cost). Capital costs are above the values that are usually calculated within RIs. Staff costs are below their usual weight within the budget. It is usually around 50% of the budget.

In terms of FTEs, in total around 7,000 FTEs participate in the operation of French RIs and partly French RIs. A small portion are directly employed by RIs. On average one FTE in the framework of French RIs costs annually €K 67. At international level, the annual cost per FTE is, on average, €K 136.

According to the same survey, the number of institutions providing staff to a single infrastructure can be as high as 30. This level of participation questions the management model of RIs when there is a high number of partners.

The survey insists on the fact that it is important to systematically consider the decommissioning of an installation and the potential costs associated. The estimated cost of decommissioning can be one to two years of operation for the most complex cases. Even if the RI has no physical assets, its dismantling costs can be associated with reclassification or termination of employment contracts. Such a cost can be reduced by the fact that a lot of staff are provided in-kind. In case the RI is ruled by private law, it could be subject to a redundancy scheme.

1.2.2. RECOLNAT results

At the time of the survey, MNHN participated with the RECOLNAT research infrastructure.[10] As the project leader, MNHN gathered the information from 32 members (out of 40). The assumption was that the perimeter of the research infrastructure was natural history collections: to provide access to these objects (and related activities) is the RI itself. It means that all activities related to Natural History collections within the 32-member institutions were considered as part of the RECOLNAT perimeter. The first result was that MNHN provided 85% of the resources of the RI. In 2017, the “full cost” of RECOLNAT MNHN was €M 32. The distribution of the costs was the following:

- 49% = amortization
- 8% = operating cost (except staff)
- 23% = staff costs
- 20% = indirect cost

[10] The National Network of Natural History Collections (RECOLNAT) is a French research infrastructure (RI). It concerns all-natural history collections and their valorisation by research. At the heart of its missions is the production and provision of a corpus of data for the study of current and past geology and biodiversity.

CHAPTER 2. Methodology implemented for DiSSCo Cost Book

DiSSCo Prepare Cost Book methodology is inspired by the costing processes presented above: from ICEDIG and the French Ministry of Research (MESRI). At the beginning of the WP4, in February 2020, DiSSCo objectives had been defined – notably to provide a unique point of access to Natural Science Collections (NSCs)– but the rules of operation of the Research Infrastructure were still unclear. With cost calculation, there is a need to clearly understand the working organisation behind the objective in order to identify the origin of costs and then calculate them. The WP4 developments towards the Cost Book reflect the developments of DiSSCo implementation along DiSSCo Prepare.

The first steps to develop the methodology were based on the perimeter of the future RI. It was based on an understanding of the activities linked to access to NSCs. These activities were structured into different large categories of costs. They correspond to the perimeter delimited above (see pages 8 and 9). On that basis, two pilot projects were implemented within WP4. Based on those tests, resource persons were identified for the different categories of costs. The work was then subdivided into three main tasks. The methodology is not exactly the same for each of these three tasks. The Central Hub Office and IT services and architecture were seen as centralised activities. In that context, a small team was contacted and met in order to define the perimeter, cost units and to make the calculation. For the National Nodes-related services, it was more complex as those costs are highly decentralised. WP4 had to develop a methodology that was accessible for as many users as possible, despite languages and cultural differences. A strategy and tools were designed in order to collect as much costing data as possible.

The following section presents how the methodology for the Cost Book was designed, the general principles followed and a presentation of its implementation. It includes recommendations on how to encourage partners – notably those who are not financial - to calculate the costs and outlines the financial rules selected for the exercise on NSCs stakeholders.

2.1 Development of the Cost Book methodology

2.1.1. Delimitation of DiSSCo perimeter

The perimeter was delimited at the very beginning of the task 4.1 (Cost Book for DiSSCo). It was the result of the MNHN understanding of the DiSSCo framework and a dialogue with National Node representatives and DiSSCo main stakeholders. This first development helped to engage the partners in the Cost Book process and to test how ready they were to calculate DiSSCo costs.

During this very early stage, it was understood that one of the major challenges for WP4 was that the DiSSCo costs were decentralised. It means that the cost calculation can mobilise hundreds of individuals across Europe.

The chosen perimeter corresponds to the one explained above. The basic understanding was that DiSSCo is about access (physical or digital) to Natural Science Collections. This access can be subdivided under different types of activities. All of these related activities / services are represented in the cost areas of the Cost Book. One of MNHN's positions was to have a large understanding of the DiSSCo perimeter: from collections conservation to IT tools (see appendix n°4).

2.1.2. Testing at institutional level

Once the perimeter was delimited, MNHN had to find a way to calculate all the costs associated. In order to test an initial methodology, a pilot project was implemented between October 2020 and January 2021 within the MNHN. It was mainly inspired from the cost methodology proposed by the MESRI in 2016 and it concerned the direct and indirect costs. Regarding direct costs, the decision was taken to meet teams from the collection departments, analytical facilities and IT department. For indirect costs, the decision was taken to meet teams from the resource department (finance, property, HR). In total, nearly 90 individuals were contacted and participated in the activity.

The results of this pilot project were that this costing exercise was:

- highly complex
- time-consuming
- required a clear delimitation of the perimeter with a very clear understanding of the RI.

The better the WP4 team knew the managers leading the different department, the faster and more complete the results were. Speaking about costs was therefore also speaking about trust. The costing process also reveals the structure of an organisation. In the case of MNHN, the finding was that, due to the lack of a unified procedure, it was complex to define a harmonised methodology to trace the costs.

Thanks to this first test, MNHN identified the information that was essential and the one that was not. **Building and improving the methodology was then about finding a balance between an acceptable level of accuracy and the amount of time required to find the information.** In some cases, the amount of work required to establish fully accurate figures meant the exercise was more expensive to carry out than the costs of the research infrastructure itself.

2.1.3. Key results from MNHN

The MNHN pilot shows how the information gathered using the methodology can be useful for an institution and for the DiSSCo RI as a whole.

2.1.3.1 Overview of the Collections department budget and staff distribution

The first element is the estimation of the resources allocated to the services that, in a broad sense, DiSSCo institutions are currently offering. It contributes to the estimation of the feasibility of any project.

Figure 3 – MNHN Collection department budget distribution – 2019

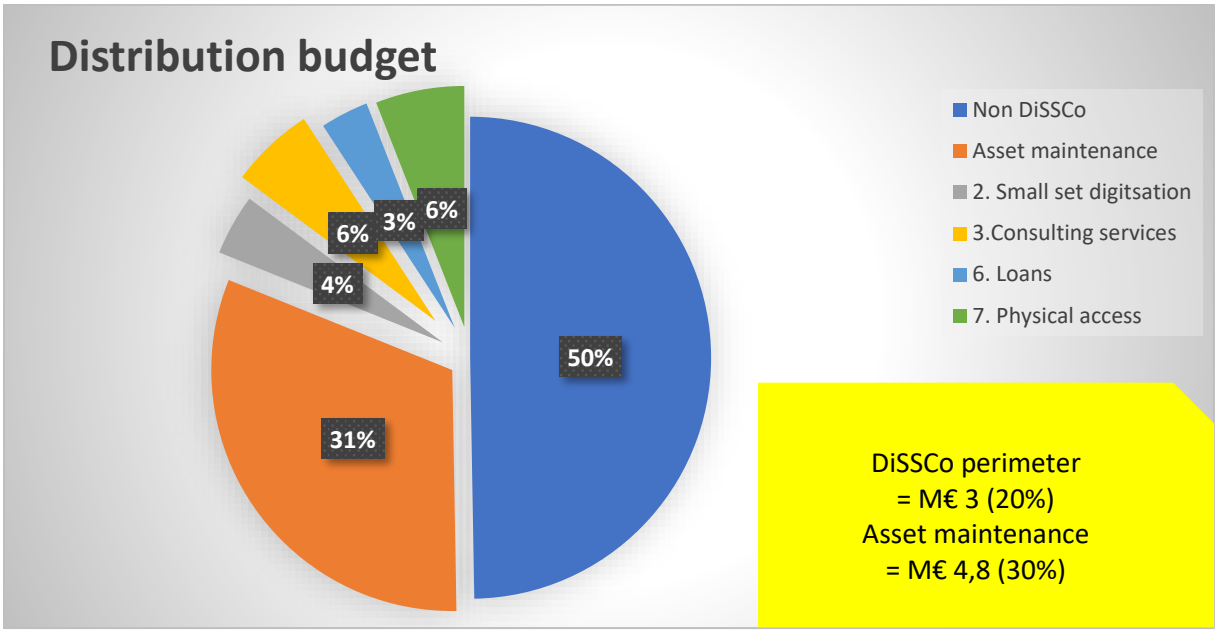
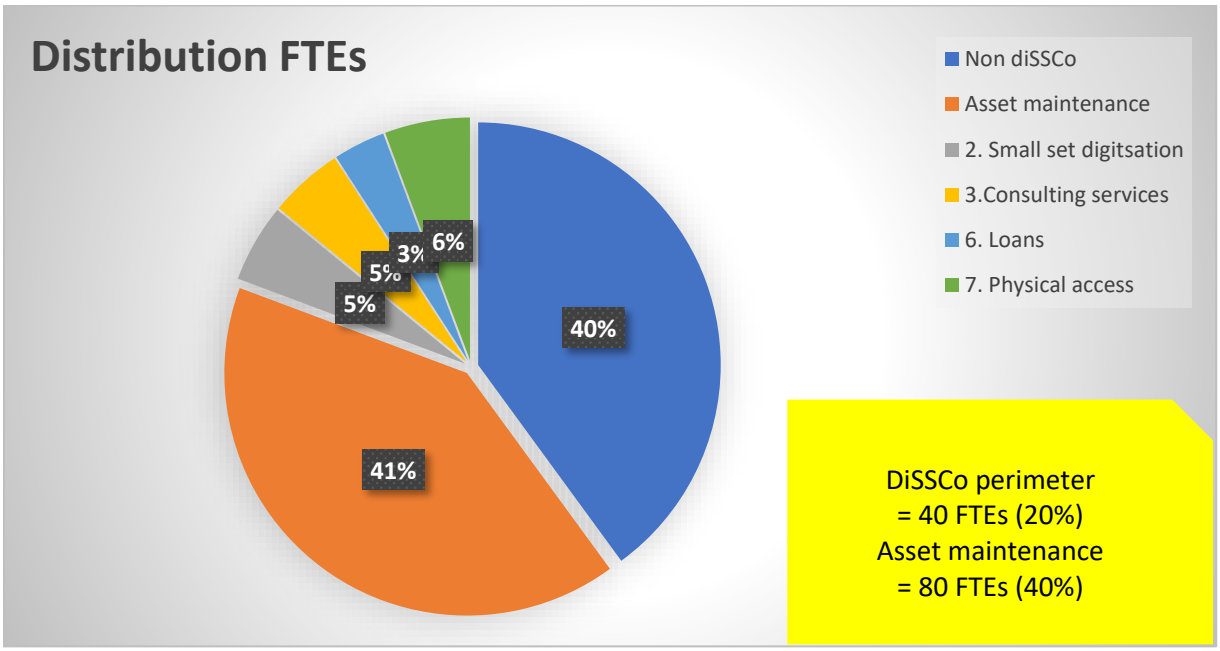


Figure 4 – MNHN Collection department staff distribution – 2019



MNHN researchers are involved in DiSSCo activities, but they are also, if not mainly, involved in research and teaching activities. This is evidenced by the two diagrams above which show that non-DiSSCo activities are significant and that these non-DiSSCo activities are undertaken by staff on higher salaries than staff involved in DISSCo activities.

2.1.3.2 Overview of the collection department types of costs

Figure 5 – MNHN collection department structure of costs – 2019

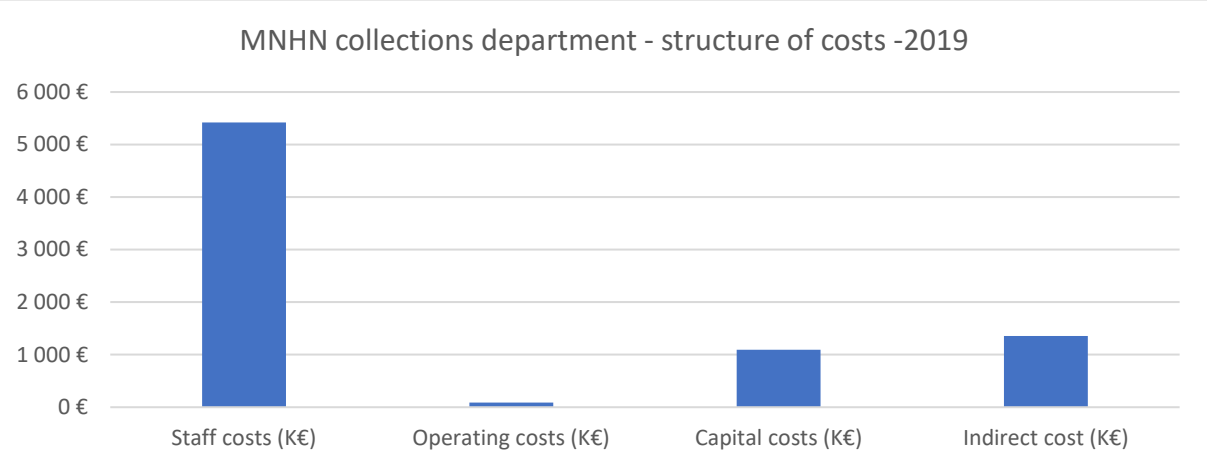
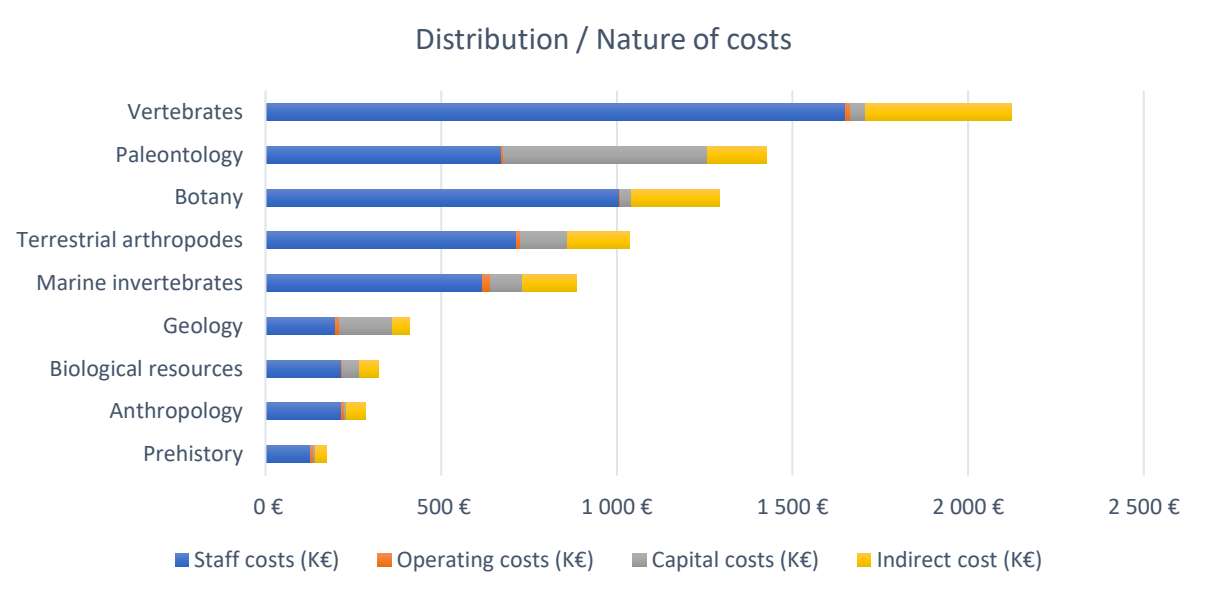


Figure 6 - MNHN collection department detailed structure of costs – 2019



The MNHN is a large institution which is structured in collection departments corresponding to the different natural history disciplines (botany, zoology…) and different analytical platforms (CT-Scan, …). Therefore, each structure has been considered as an accounting entity. Whatever the size of these entities, within the collections department there is no significant difference regarding the nature of costs: staff costs are the main source of the costs; the other main component is the indirect, support costs taken as 17% of the staff costs. The method to assess these indirect costs could not be implemented at the MNHN for lack of relevant data.

2.1.3.3. Integration of conservation costs as indirect costs

Maintaining the collections in good condition is a prerequisite for offering access services and the costs incurred by this activity should be accounted for in a full cost analysis. It is difficult to pinpoint the nature of these costs in a Cost Book.

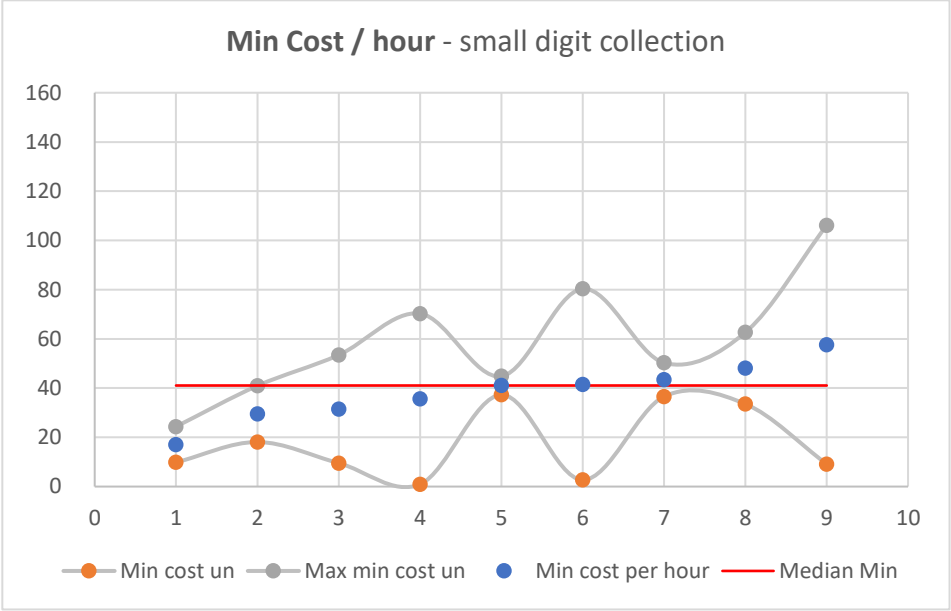
In order to estimate the importance of the preservation costs, two values were calculated:

- minimum operating costs directly connected to the DiSSCo services;
- maximum operating costs by adding the preservation costs to those costs.

Figure 7 – Cost per hour of small set digitisation / MNHN - 2019

Dots in grey are the maximum costs according to the uncertainty calculation, the ones in orange correspond to the minimum according to the same calculation.

1	Geology
2	Botany
3	Prehistory
4	Anthropology
5	Marine invertebrates
6	Biological resources
7	Terrestrial arthropodes
8	Vertebrates
9	Paleontology



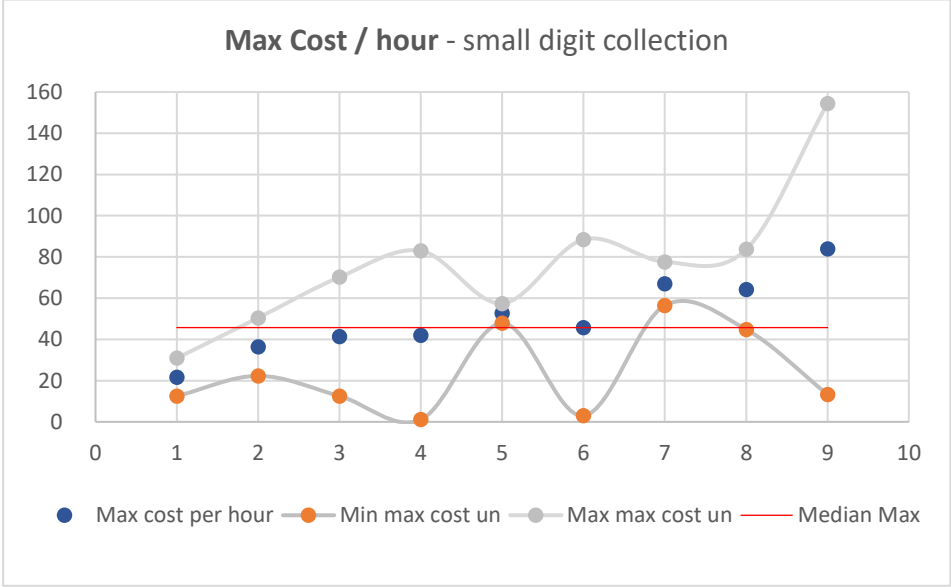
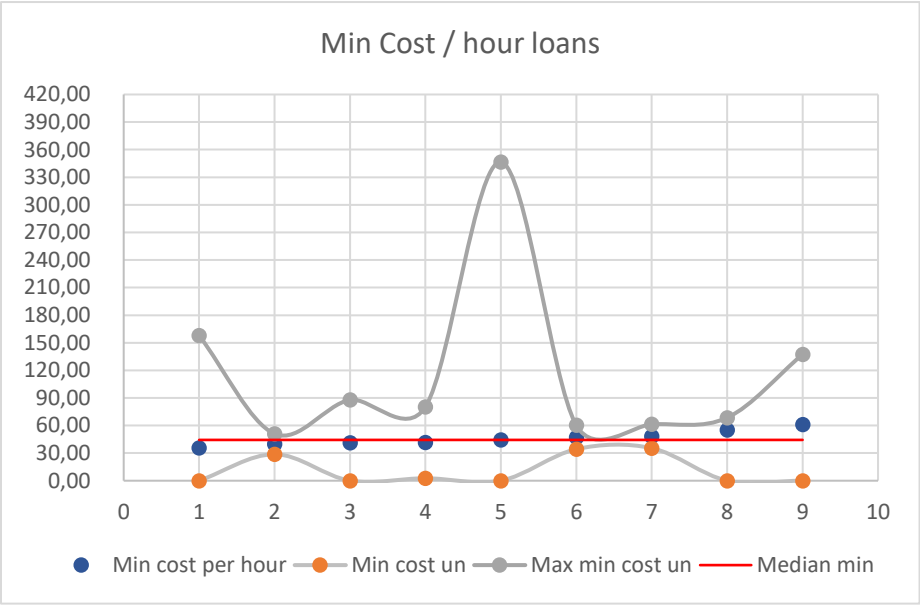
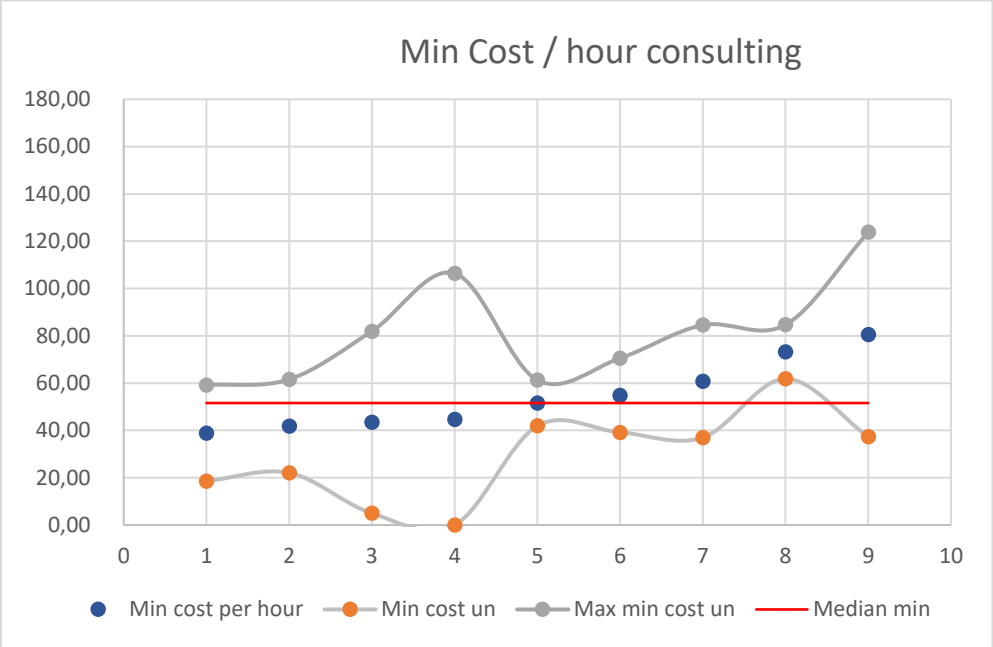


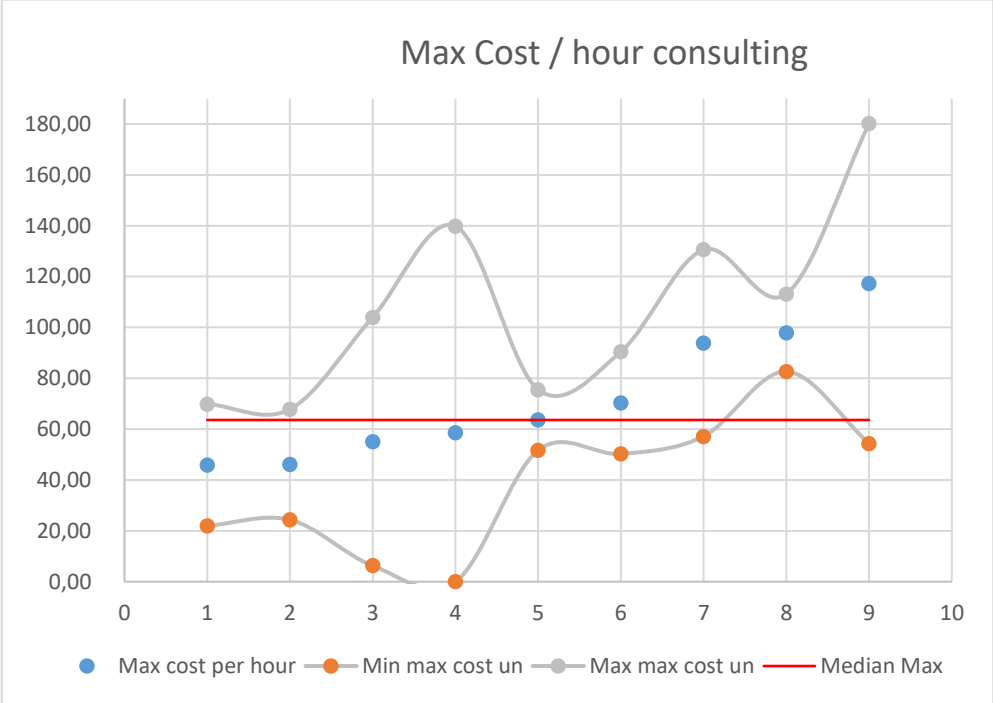
Figure 8 – Cost per hour of loans provision / MNHN - 2019



1	Prehistory
2	Botany
3	Geology
4	Biological resources
5	Anthropology
6	Terrestrial arthropodes
7	Marine invertebrates
8	Vertebrates
9	Paleontology

Figure 9 – Cost per hour – consulting services / MNHN - 2019





1	Anthropology
2	Biological resources
3	Geology
4	Prehistory
5	Botany
6	Marine invertebrates
7	Terrestrial arthropodes
8	Vertebrates
9	Paleontology

Comparing the cost/hour for the different services and collection department leads to two conclusions:

- The uncertainty is variable for each department; this is related to the number of staffs in the department: the larger the number, the more precise the estimation. The large number is not only influenced by the number of staffs, but also by exceptional circumstances; the palaeontology department was involved in a major renovation project during the reference year of the study (2019) and therefore the time spent on the DiSSCo services was very limited, hence a small number of FTEs for the services and a large uncertainty.

- The cost/hour for digitisation, loans and physical access is about the same for each department. This shows that the compositions of the teams are homogeneous, whereas they are different for consulting services. This is because staff with expertise and/or knowledge are generally situated within the top salary range which tends to be higher and larger than the low range.

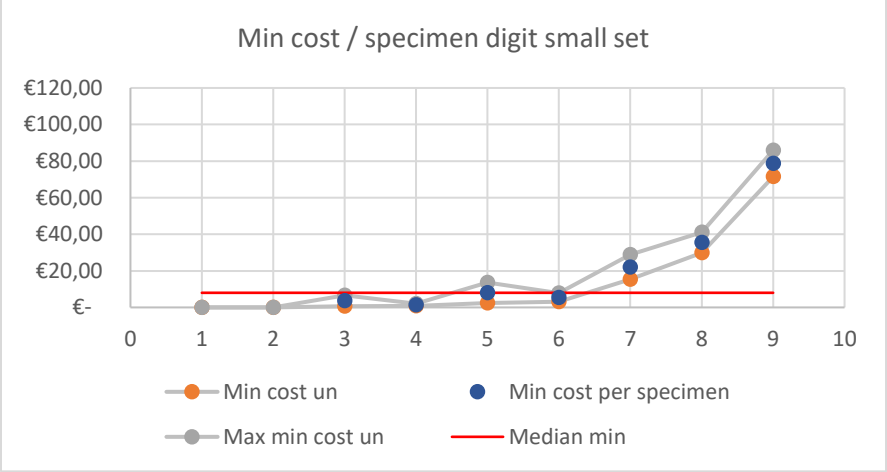
Considering that the cost/hour for digitisation, loans and physical access is about the same, it is possible to derive some conclusions from the cost/specimen for each service and department.

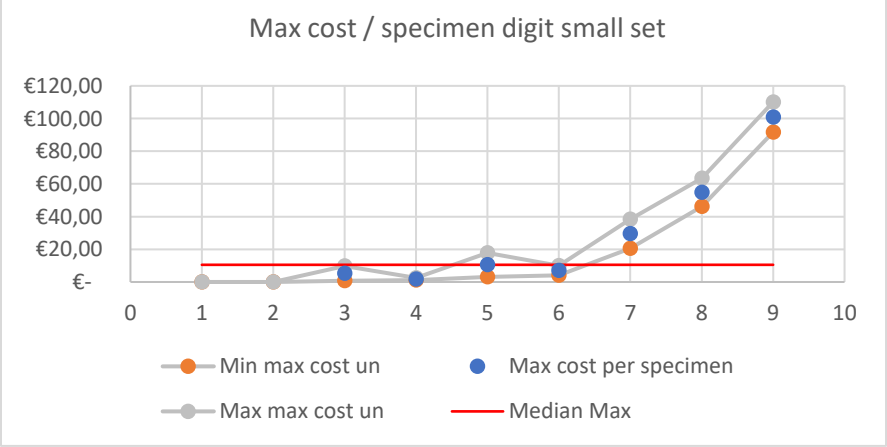
2.1.3.4. Analysis of MNHN cost per service (digitisation, loans, consulting, visit)

There are significant differences between the collection departments, which could reflect either structural differences or a bias created by different practices and organisations.

Small set digitisation: there are probably structural differences between the collections as the handling of specimens of vertebrates, insects and marine invertebrates is different from, for example, botany. However, there is also a bias if comparing marine invertebrates and insects for which handling procedures are similar. This bias probably stems from different MIDS levels of the databasing adopted during digitisation.

Figure 10 – Cost per digitised specimen (small set digitisation) / MNHN - 2019

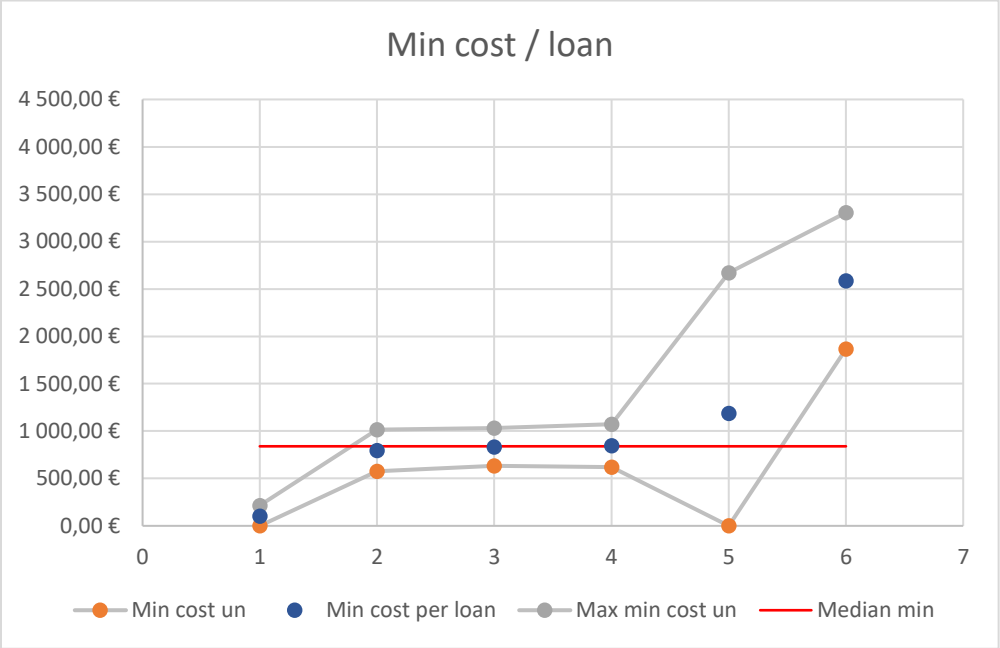


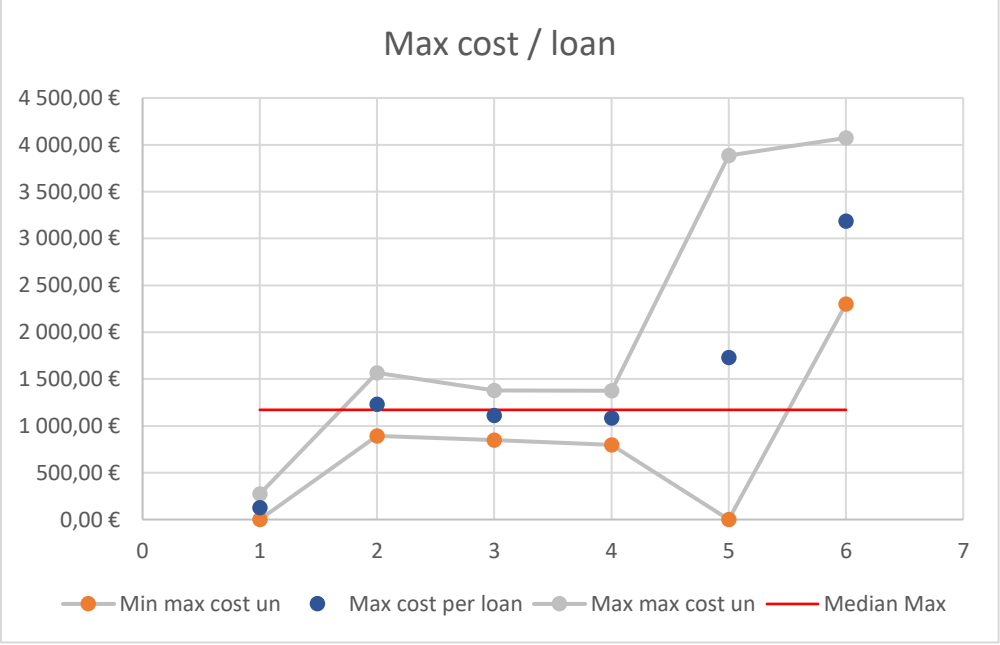


1	Anthropology
2	Biological resources
3	Paleontology
4	Botany
5	Prehistory
6	Geology
7	Vertebrates
8	Terrestrial arthropodes
9	Marine invertebrates

Loans: low value for geology and high value for botany would require examining the procedures used by these two services. There are structural reasons: size, access to the collection, database maturity. For botany, there could also be a bias in recording the information due to databasing outgoing loans and returned loans (updating the information).

Figure 11 – Cost per loan / MNHN - 2019



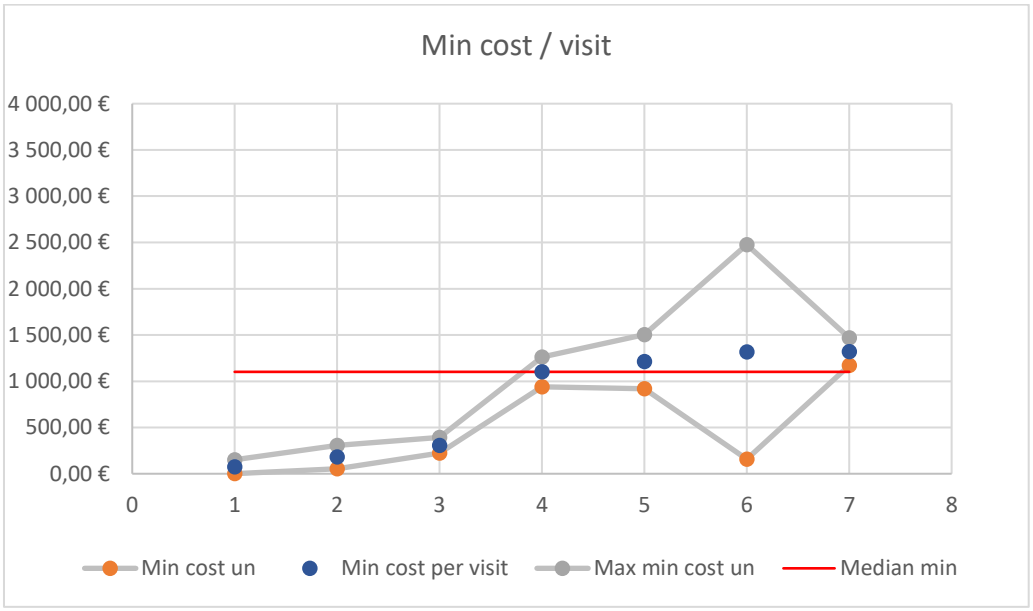


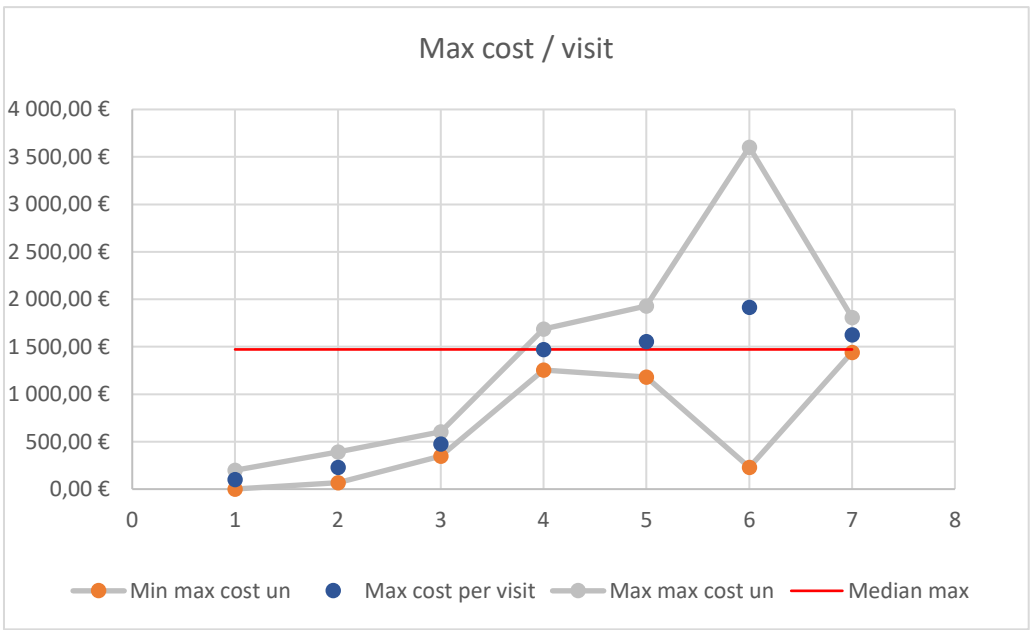
1	Geology
2	Terrestrial arthropodes
3	Vertebrates
4	Marine invertebrates
5	Paleontology
6	Botany

Visits: there is a clear and very significant difference between two groups of departments; this difference is likely due to the duration of the visits and if the collection staff are required to accompany and provide support to the visitors.

The conclusion from this section is that a unique cost/hour can be estimated for basic services for the institution, but the distinction between collections departments must be made. There is no unique cost/specimen given the uncertainty on the cost data.

Figure 12 – Cost per physical visit / MNHN - 2019



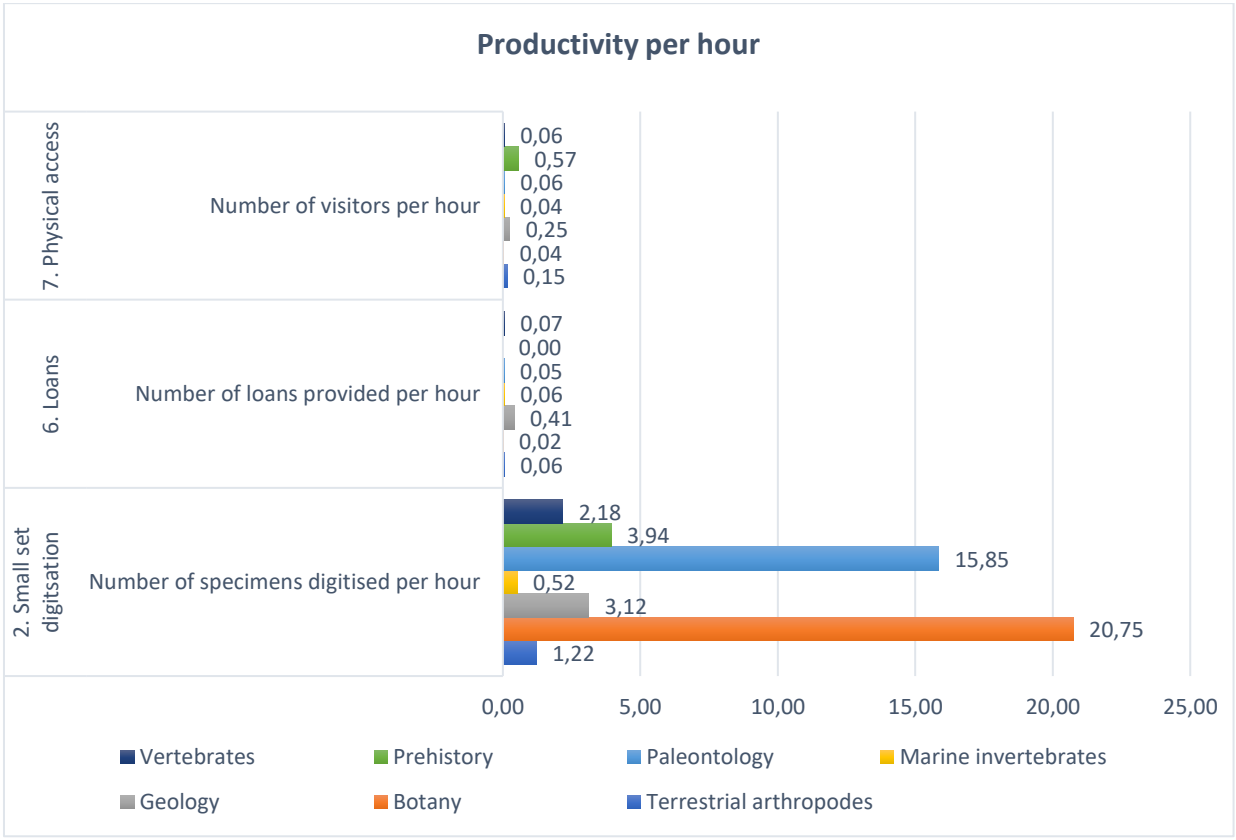


1	Prehistory
2	Geology
3	Terrestrial arthropodes
4	Vertebrates
5	Marine invertebrates
6	Paleontology
7	Botany

2.1.3.5. MNHN Collections department productivity data

The specificity of the Collections department is also visible in the productivity data (number of specimens processed per hour).

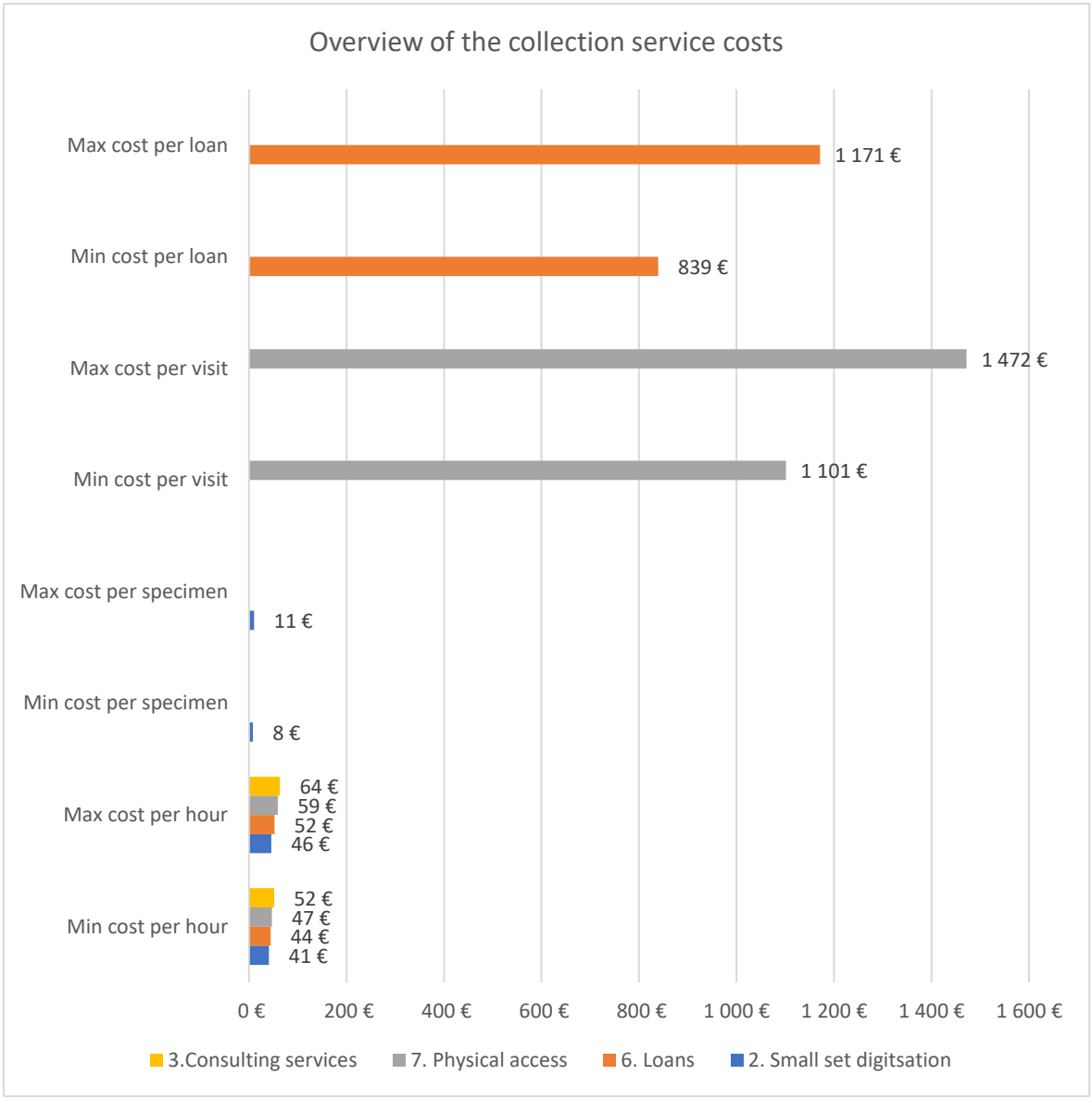
Figure 13 – Comparison cost areas and productivity per hour



The differences are significant; they are due to the nature of the collections. Knowing the cost/hour and the cost/specimen allows for assessing the feasibility of a project, that is whether it is possible/acceptable in terms of duration and budget. Even if the Collections department level is the most accurate to account for the cost of the services, a global comparison of services across the institution is still relevant.

Physical access and loans are the most significant part of the Collections department's activities. They are also the costliest services provided by the institution. The data gathered during the pilot study may be biased, especially if digitisation/databasing is part of the loan or visit process. However, the difference in cost is probably still very significant and vindicates the DiSSCo project and aim. Digitisation of collections is a good option as it will save time; it will help in preparing access or implementing remote working. A digitisation program will save resources in the institutions which will be deployed to document and provide better support to visitors, who will have better prepared their visits.

Figure 14 – Overview – cost calculated / MNHN - 2019



2.1.3.6. MNHN analytical facilities results

Figure 15 – MNHN analytical facilities – structure of costs – 2019

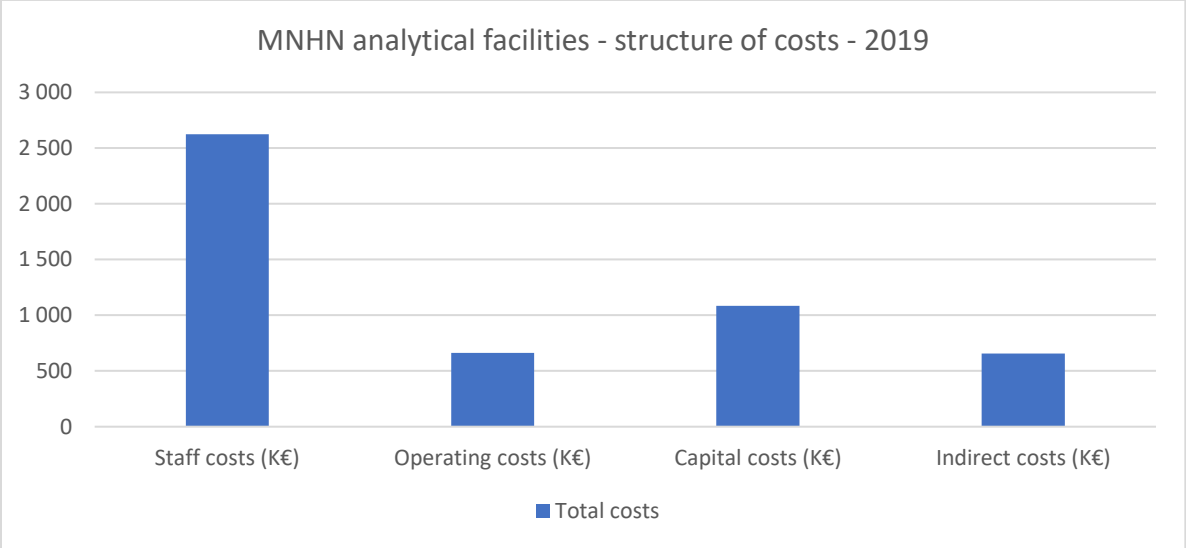
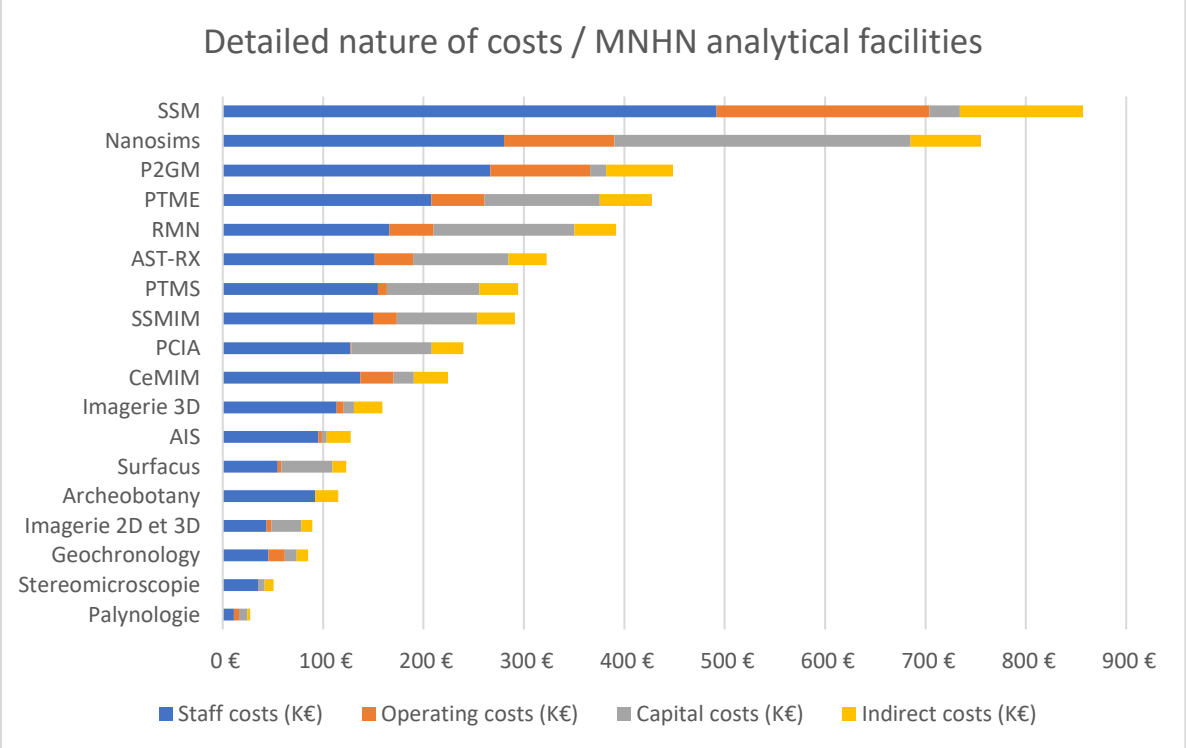


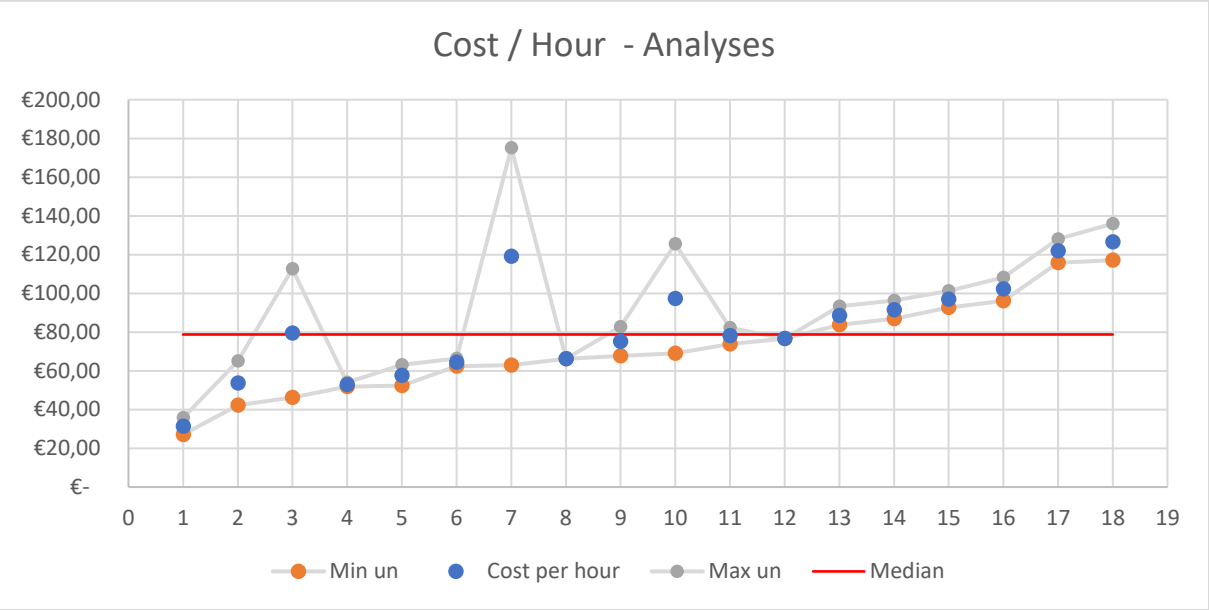
Figure 16 – MNHN analytical facilities – detailed structure of costs – 2019



The analytical facilities at MNHN are more diverse in nature and practices than the collection departments. Hence the difficulty, if not the impossibility, to fix a unique rate for basic costs. Costs are specific to facilities and attempting to standardise costs have to be undertaken at the whole RI level.

Figure 17 – Cost per hour of analytical facilities / MNHN - 2019

The acronyms listed (AIS, SSM, SSMIM, etc.) correspond to the names of the analytical facilities of the MNHN.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
AIS	Archeobotany	Geochronology	SSM	Imagerie 3D	P2GM	Palynologie	Stereomicroscopie	CeMIM	Imagerie 2D et 3D	AST-RX	Surfacus	SSMIM	PTMS	PTME	PCIA	RMN	Nanosims

2.1.4. Test at European level

Another challenge of the Cost Book methodology was to make sure that the concepts used were accessible for European partners. The fact that DiSSCo stakeholders speak different languages, have different accounting systems and different work practices introduces great risk for the methodology as, in order to have coherent results, a shared understanding is a prerequisite.

The decision was taken to share a new version of the methodology with the WP4 partners (NHM, RBINS, APM, SGN).

This second pilot project with WP4 members was a way to test if it was possible to have a shared understanding of the methodology, despite administrative, financial, and language differences. All the partners completed the table and sent their feedback to MNHN. A meeting was organised with MNHN in order to get their direct feedback and exchange on how to simplify the methodology. Their general feedback was that the exercise was very difficult and time consuming. The perimeter was still not fully clear.

From the MNHN side, the results of this survey were also that the information received was not easy to use. It was providing large budget masses. The basis of the DiSSCo Cost Book was to understand the costs in order to then potentially develop tariffs based on the effort implemented by DiSSCo partners.

Following this second pilot project, **the objective was not only to collect the cost distribution but also to ask for the number of services provided per year (number of loans, number of visits, number of specimens digitised, etc.). With that information, it is possible to calculate the production cost of the services.**

2.2 General methodology requirements

In order to understand the methodology implemented by MNHN, some general rules can be listed. They guided the implementation process and facilitated work with a large number of partners, who are not always fully integrated in the DiSSCo process.

2.2.1. Soft skills requirements

- **Identify contact persons**

Information on costs can be in the hands of a few contact persons in position to have a good understanding of the staff activities, annual operating and investment costs. Generally, it is the managers and directors of departments. It can also be administrative managers, with information on the department costs. In order to avoid contacting too many people, it is important to identify individuals who can be pivotal in the Cost Book process.

- **Clearly define the perimeter and the types of costs to be considered**

Often, contact persons raise questions on expenditure and if it should be considered in the process. In this case, as it is not feasible to have an exhaustive list on the costs of a project implemented across 23 countries, within 170 different institutions, one solution is **to define priorities and consider that the smaller the negative impact a decision might have on the final cost calculation, the more it is possible to exclude it, instead of wasting time.**

In the case of the DiSSCo perimeter, the difficulty is that it is transversal across different activities of the institutions. It cannot be copied on a pre-existing perimeter. In that context, traditional accounting systems, with pre-existing cost categories, cannot provide all the information needed. It requires human beings who will look for the information and analyse it in accordance with the predefined scope.

- **Define an option to keep the data coherent and the methodology flexible**

In order to make sure that the data collected are coherent, there is a need to define coherent categories. The concept of accounting entities was determined: “a structure that operates some if not all of DiSSCo activities with a perimeter that respects the resources allocated (staff, expenses). It is an entity based on the principle of subsidiarity (subsidiarity is a principle of social organisation that states that social and political issues should be dealt with at the most immediate (or local) level that is consistent with their resolution). It is the organisational unit in your institution that is most able to measure the costs of a specific DiSSCo area or scientific domain.”

Such a concept allows for flexibility. WP4 does not enforce categories such as “botany, vertebrates, etc.”, it instead gives institutions the possibility to adapt the methodology to their own structure and organisation. The resource persons can then be the heads of the different accounting entities of the organisation.

- **Fluidly provide easy-to-understand documents**

Thanks to the two pilot projects, the finding was that most of the time, people do not read the documents and even the emails sent. The methodology should be short, without too much text, and if possible with images and diagrams. It allows for a rapid understanding of the concepts used and, in case there is a question, people can contact the team.

- **Hold the line: continuity and persistence**

In order to get the same data from all partners, the methodology shall not be changed over the course of implementation. In the case it would be changed, it could change the structure of data collected and the results would be more complex to process.

Another structural behaviour to adopt is to regularly send reminders. If the methodology is sent only once, it is possible that nobody will follow it. In order to be credible, it is important to repeat the request and therefore to have rather long deadlines.

- **Leave room for uncertainty**

Each time the question of costs according to a perimeter for a RI which does not exist yet is raised, individuals respond with uncertainty. The perimeter is still not fully clear. The costs are not always clearly trackable. Teams change and past practices and information are not always easily obtained. All of those variables lower the accuracy of the data gathered. From a calculation point of view, WP4 decided to define 4 levels of uncertainty: they allow people participating in the survey to self-assess the level of uncertainty of their responses. These levels are directly connected to formulas which can calculate a range between which the accurate costs are.

This method helps participants to answer even if they are not sure of their responses, and the WP4 team to associate figures with this uncertainty.

- **Define a clear objective**

At the beginning of the methodology, the objective was unclear. It was about calculating the full costs related to Natural Science Collections in Europe. The outcomes of this objective can have an impact on assessing the economic value of NSC in Europe and then during conversations with funders. It would serve as leverage to negotiate the ERIC budget insofar as it might demonstrate that integrating NSC would not be disproportionately costly in comparison to the actual cost of preserving and providing access to collections.

This objective was good but not actionable enough as an argument to implement the methodology. The second objective, which was also an incentive to implement the methodology,

was to estimate the costs paid to provide services. This would allow partners to have a better understanding of their own costs and, and if needed, be able to charge for some of their services.

2.2.2. Technical costing perimeter

The objective of the full cost methodology is to evaluate the direct and indirect costs of the projects. Direct costs correspond to staff costs, operating costs and investments directly connected to the project. Indirect costs are costs that are essential to the proper functioning of organisations and that are common to all services.

Direct costs:

The conclusions of the two pilot projects are that **staff costs** are the most important type of costs in the DiSSCo context. It means that in the Cost Book process, these are the types of costs that WP4 should focus on. At the same time, staff costs are the types of costs which might be the most complex to calculate as they require interviewing managers on the distribution of time of their teams. It also requires access to the wages of all the people involved in the DiSSCo perimeter. It is also connected to the type of contracts of the employee: part time or full time. All of this information makes staff costs essential and very complex to calculate.

Another assumption of the pilot projects was that **operating costs other than staff costs** have a minor impact on the total costs. They are not always clearly identified. In the case of DiSSCo, it is possible to use large figures and a rough distribution within the different DiSSCo cost areas. It allows us to save time and effort on data which will not have a huge impact on the final results.

Investments **/ capital costs** are a bit more complex to investigate. According to the ESFRI methodology, amortization cannot be considered in the cost calculation process. It means that only the initial cost should be considered. Such an estimation is coherent for the "centralised costs" of the RI: the hub and the IT infrastructure. It is more difficult when it comes to the institution's contribution. For this calculation, as the end result is a cost per hour, the decision was taken to ask for the total investment made during one year. It is a very unstable solution as, depending on any given year, there may have been major investment or none at all. It creates a high level of uncertainty and potential for great errors in the cost calculation process. It can have a major impact on the cost per hour.

In the meantime, this decision was taken because through the pilot projects, the finding was that the investment recording was often missing. It means that the team did not have a clear vision of past investment: purchase cost, year of acquisition, estimated lifetime. Without this information, it is complex to make the calculation of the annual amortization.

Finally, one option could be to arbitrarily decide to use an average of the investments of the 5 years preceding the cost assessment. This would capture the variability of investment, including no investment in some years and high investment in others.

Indirect costs:

Indirect costs are the costs covered by the institution that are involved in the effective functioning of projects and common to the whole institution. They can be subdivided into two categories:

- related to staff: human resources, legal services, finance department, other shared services that involve institution staff;
- related to buildings: renovation/maintenance, security, housekeeping, utilities (electricity, gas, water).

At the very beginning of WP4, the aim was to calculate exactly these costs in order to find a flat rate that was tailor-made for Natural Science Collections. During the pilot project within MNHN, different resource teams were contacted and met in order to collectively find a way to calculate these costs. The results showed that the data required for an accurate calculation of these costs was missing. For instance, it required an estimation of power consumption per room used. This information requires investment: namely specific equipment to take the measurements which is not currently available at MNHN. It was therefore more expensive to precisely calculate these costs than the potential revenue it could bring.

The decision was taken to use the flat rate proposed by the EU for these costs: 25% on direct costs. Institutions using the proposed methodology have the possibility to use their own flat rate based on their own calculation of their indirect costs.

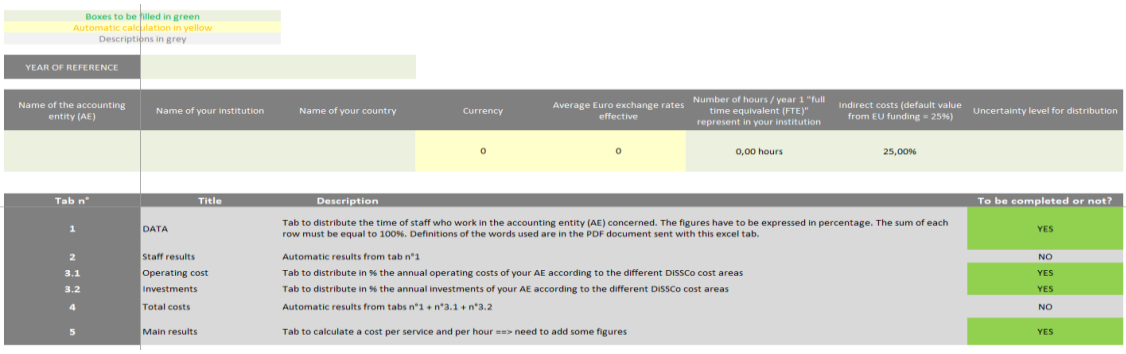
2.3 Methodology for decentralised costs: National Nodes

The methodology is composed of one document explaining the concepts used and two Excel tables with similar structures. The package also contains two examples from MNHN full costs.

2.3.1. General information with automatic impact

The first Excel table serves as a tool to guide the partners towards the results. It first asks them **general information about their institution** and notably the yearly hours of work per full time equivalent (FTE). Once participants enter the name of their country in this tab, the exchange rate is automatically calculated. Finally, in this first tab, the partners are asked to self-assess their level of uncertainty when they are completing the table. Level 1 means that they are completely sure of their response. Level 2 means that their responses have a 5% margin of error, level 3 means 10% margin of error and level 4, a 25% margin of error. All of these responses have an automatic impact on the other tabs of the table.

Figure 18 - Tab 1 – Cost assessment tool



Boxes to be filled in green
Automatic calculation in yellow
Descriptions in grey

YEAR OF REFERENCE	

Name of the accounting entity (AE)	Name of your institution	Name of your country	Currency	Average Euro exchange rates effective	Number of hours / year 1 "full time equivalent (FTE)" represent in your institution	Indirect costs (default value from EU funding = 25%)	Uncertainty level for distribution
			0	0	0,00 hours	25,00%	

Tab n°	Title	Description	To be completed or not?
1	DATA	Tab to distribute the time of staff who work in the accounting entity (AE) concerned. The figures have to be expressed in percentage. The sum of each row must be equal to 100%. Definitions of the words used are in the PDF document sent with this excel tab.	YES
2	Staff results	Automatic results from tab n°1	NO
3.1	Operating cost	Tab to distribute in % the annual operating costs of your AE according to the different DiSSCo cost areas	YES
3.2	Investments	Tab to distribute in % the annual investments of your AE according to the different DiSSCo cost areas	YES
4	Total costs	Automatic results from tabs n°1 + n°3.1 + n°3.2	NO
5	Main results	Tab to calculate a cost per service and per hour ==> need to add some figures	YES

2.3.2. Distribution of staff time

The second step of the methodology is to distribute within each accounting entity of the institutions the **time spent by the staff on the different categories of activity** that are in the DiSSCo perimeter (digitisation, consulting services, analysis, loans, physical access). In addition, a category concerning "asset maintenance" was added. It includes the time spent to preserve the collections and to "curate the collections' data" (manually and computationally). The sum of these cost areas represents the costs that can be associated with DiSSCo. In addition, there is a section called "non DiSSCo" which is the subtraction between DiSSCo costs and other costs. The sum of it should be equal to 100% of each FTE of the accounting entity. These FTEs are linked to their full economic cost (FEC) which is equal to the total cost of the employee for the employer. The result of this aggregation of information is the annual staff costs of the accounting entity and the distribution of this annual cost within the different cost areas.

Figure 19 - Tab.2 - Staff cost distribution

Distribution of staff working hours

Name of the accounting entity	0

Identification (ID) (1 row = 1 person)	Full time equivalent (FTE)	Full economic cost (FEC) per year in your currency	non DiSSCo	Mass digitisation	Small set digitisation	Consulting services	Training	Analysis	Loans of collections	Physical visit to collections	Asset maintenance expenditure = preservation / conservation of collections + data curation	Total	Check
#1		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#2		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#3		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#4		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#5		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#6		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#7		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#8		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#9		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#10		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#11		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#12		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#13		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#14		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#15		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#16		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#17		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#18		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#19		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#20		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#21		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#22		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#23		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#24		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#25		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#26		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#27		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#28		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX
#29		0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	FAUX

2.3.3. Operating and capital costs

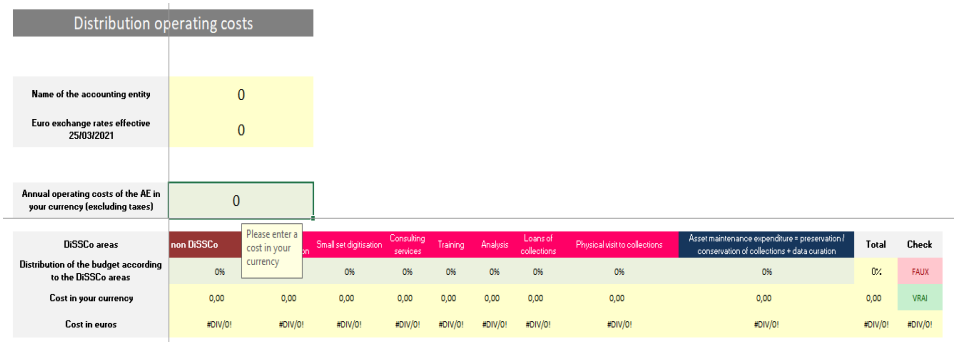
The next two steps are related to operating costsand capital costs.[11] Here it is understood that WP4 needs rough estimations as, according to the first pilot projects, **such costs are less important compared to staff costs**.[12]

In this respect, partners were asked to add a general annual cost for both categories and to distribute them in % among the exact same categories as for the staff costs.

[11] See appendix 1, financial glossary

[12] This conclusion was made on the basis of MNHN data and the understanding was that there was less investment and more staff costs. This assumption can be discussed and more financial data could clarify it.

Figure 20- Tab 3- Operating costs: same structure for investments

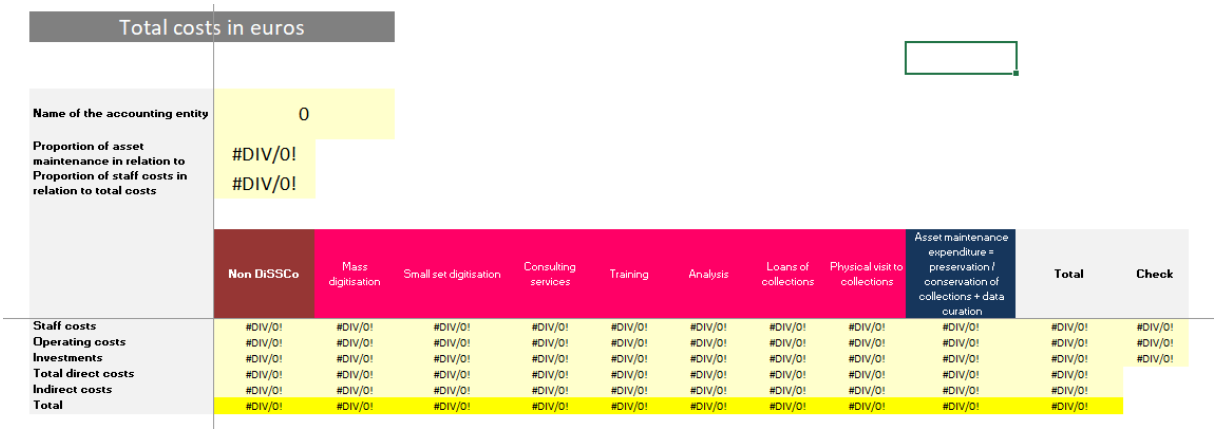


2.3.4. Aggregated first results

After these three steps, some initial general costs are calculated:

- The proportion of asset maintenance in relation to total costs: in the accounting entity, a calculation of the weight of natural science preservation costs in relation to the other activities implemented within the AE;
- Proportion of staff costs in relation to total costs;
- The total costs of each DiSSCo area implemented within the AE;
- The proportion of non-DiSSCo costs in relation to the total costs;
- An estimation of the indirect costs: here the solution proposed to use the 25% that is implemented for European projects. The 25% rate is linked to staff costs as sometimes operating costs and investments can be outsourced – which cannot be linked to indirect costs.

Figure 21 - tab 4 - Automatic calculation, total costs



2.3.5 Calculation of cost per unit (hour, specimen, loan, etc.)

The last tab of the same table summarises all the automatic calculations made based on all the data gathered through the table. It is subdivided among the different DiSSCo cost areas and repetitively lists the costs calculated. In relation to the level of uncertainty registered within the first tab, each estimated cost is here associated with its uncertainty. It is materialised with a figure in euros. The range between the low and high levels of uncertainty is where the exact figure should be located.

For the different cost areas, some additional information is required:

- Number of specimens digitised per year (for mass and small-set digitisation);
- Number of images produced per year (for mass and small-set digitisation);
- Number of demands for consulting services completed per year;
- Number of training courses provided per year;
- Number of loans provided per year;
- Number of specimens loaned per year;
- Number of visitors per year;
- Number of visitor days per year.

With this additional information, it is possible to calculate a cost per specimen digitised (according to the type of digitisation), a cost per image produced (according to the type of digitisation), a cost per training, a cost per loan, a cost per specimen, a cost per visitor and a cost per visitor day. It allows for a comparative approach among DiSSCo member institutions.

For each cost per category, there is a minimum cost and a maximum cost. The minimum cost corresponds to the aggregation of all the costs which allows the service to be provided and its division by the number of actions done (listed above). The maximum cost corresponds to the aggregation of all the costs provided and the asset maintenance proportional to its weight in relation to the AE total costs. This option was chosen in order to value the maintenance / preservation of natural history collections. It can be understood that without this maintenance, DiSSCo services would not exist as the collections would no longer be accessible. It is an option for institutions to add the annual costs to preserve and maintain the collections to the formulation of their costs.

In addition to that information, the data gathered can provide large data on the collections and services related. With all of that data, it is possible to estimate the number of FTEs involved in the work on Natural Science Collections at European level. We can have an estimation of the economic value of these activities. It is therefore possible to know what kind of costs are linked to the collections.

Figure 22 – Tab 5 – Cost assessment tool, main results

Results to copy and paste in the tab 'institutional costs for services' - tab 2.1

	Type of costs	DATA	Check
	Detailed name of the AE	0	
	Total costs of the AE	#DIV/0!	#DIV/0!
	Proportion of staff costs in relation to total costs	#DIV/0!	#DIV/0!
	Total asset maintenance expenditure in the AE	#DIV/0!	#DIV/0!
	Proportion of asset maintenance	#DIV/0!	#DIV/0!
	Total FTEs	0,00 FTE	VRAI
1. Mass digitisation	Total costs per year	#DIV/0!	#DIV/0!
1. Mass digitisation	Additional cost with asset maintenance	#DIV/0!	#DIV/0!
1. Mass digitisation	Total FTEs	0,00 FTE	VRAI
1. Mass digitisation	Total hours per year	0 hours	VRAI
1. Mass digitisation	Min cost per hour	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Min cost per hour)	#DIV/0!	
1. Mass digitisation	Max cost per hour	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Max cost per hour)	#DIV/0!	
1. Mass digitisation	Number of specimens digitised per year	0	
1. Mass digitisation	Number of specimens per hour	#DIV/0!	#DIV/0!
1. Mass digitisation	Min cost per specimen	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Min cost per specimen)	#DIV/0!	
1. Mass digitisation	Max cost per specimen	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Max cost per specimen)	#DIV/0!	
1. Mass digitisation	Number of images produced per year	0	
1. Mass digitisation	Number of images produced per hour	#DIV/0!	#DIV/0!
1. Mass digitisation	Min cost per image	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Min cost per image)	#DIV/0!	
1. Mass digitisation	Max cost per image	#DIV/0!	#DIV/0!
1. Mass digitisation	Uncertainty +/- (Max cost per image)	#DIV/0!	
2. Small set digitsation	Total costs per year	#DIV/0!	#DIV/0!
2. Small set digitsation	Additional cost with asset maintenance	#DIV/0!	#DIV/0!
2. Small set digitsation	Total FTEs	0,00 FTE	VRAI
2. Small set digitsation	Total hours per year	0 hours	VRAI
2. Small set digitsation	Min cost per hour	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Min cost per hour)	#DIV/0!	
2. Small set digitsation	Max cost per hour	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Max cost per hour)	#DIV/0!	
2. Small set digitsation	Number of specimens digitised per year	0	
2. Small set digitsation	Number of specimens digitised per hour	#DIV/0!	#DIV/0!
2. Small set digitsation	Min cost per specimen	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Min cost per specimen)	#DIV/0!	
2. Small set digitsation	Max cost per specimen	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Max cost per specimen)	#DIV/0!	
2. Small set digitsation	Number of images produced per year	0	
2. Small set digitsation	Number of images produced per hour	#DIV/0!	#DIV/0!
2. Small set digitsation	Min cost per image	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Min cost per image)	#DIV/0!	
2. Small set digitsation	Max cost per image	#DIV/0!	#DIV/0!
2. Small set digitsation	Uncertainty +/- (Max cost per image)	#DIV/0!	
3.Consulting services	Total costs per year	#DIV/0!	#DIV/0!

2.4 Methodology for centralised costs: Central Hub Office and IT Team

At the beginning of WP4, the hub and the IT infrastructure of the research infrastructure were merged with the calculation designed for the nodes. At the time, it was still unclear how DiSSCo RI was going to be managed. The decision that DiSSCo was going to be a European Research Infrastructure Consortium (ERIC) was taken in 2022, in other words two years after the beginning of WP4. The more WP4 understood how the Research infrastructure is expected to work, the more WP4 identified the people to be contacted in order to estimate costs.

With the cost calculation exercise, the objective determined by the DiSSCo Coordination and Support Office (DiSSCo CSO) **was to estimate the minimum budget in order to allow the implementation of the research infrastructure.** The assumption is that the RI has to start with the minimum viable budget and grow over time. Once in operation, DiSSCo Central Hub can develop an enlargement strategy, respond to calls for proposals, diversify the source of its funding and increase the budget for DiSSCo programs.

The calculation methodology for the Central Hub is based on cost units: number of FTEs, number of days, number of packages associated with different costs per unit. It means that if the number of cost units required evolves, the DiSSCo budget can be adjusted and reflect the efforts needed to run the RI.

When it comes to cost calculation, **the location of the teams can have an impact on the figures calculated:** inflation is different from country to country. Not all DiSSCo partners are using euros as

national currency and therefore cost estimations are subject to changing currency exchange rates. Wages are different for developing an identical service from country to country. All of these parameters mainly lead to assumptions and not to definitive values.

In that framework, one option is to work on costs from an identified country and if there is a change of location, it is possible to use a "country correction coefficient (CCC)" (see appendix n°10). The use of this tool allows us to adjust the figures calculated proportionally to the price level of the countries. It allows working hypotheses and leaves the door open for changes in the future.

The current working assumption, based on the current location of DiSSCo Coordination and Support Office (CSO) is that the hub will be hosted by the Netherlands. This hypothesis can be adapted if another decision is taken.

Regarding the IT infrastructure and e-services, the information available in order to calculate their costs is, to some extent, decentralised. The funding to develop e-services came mainly from EU projects which were divided into several work packages (WP). The same WPs are currently led by different institutions, located in different countries.

This manner of working has led to the development of three main working hypotheses on the future DiSSCo services:

1. Institutions (or a consortium of several institutions considered as "units") located in different countries could build and implement DiSSCo digital services (see figure 2). It means that there would be SLAs with the institutions that would become service providers for the RI. SLAs are contracts that describe the service provided and the expectations regarding the implementation of the services. In terms of costs, an assumption is that within these contracts, there would be an agreement on an annual budget provided by the ERIC to the institutions to fund the implementation of the service.
2. The Central Hub could regroup all the IT infrastructure and the e-services associated. The costs would then be centralised.
3. Hybrid option: a large part of the services would be covered by the hub and a minority of them covered by DiSSCo members.

As contact persons were not the same for the core and for the centralised DiSSCo activities, a decision was taken to split this work into two main steps:

1. The costs to run the Central Hub Office
2. The costs to build and operate the Central Hub digital tools and services.

2.4.1. Central hub Office cost calculation

The Central hub Office cost calculation has been done in partnership with the DiSSCo strategy team. In the first place, the team defined the roles needed to run the RI. These roles were then associated with job positions: people who will form the RI central office team and coordinate its actions. Finally, wages were calculated on the basis of European Commission salary scales. To calculate the other costs linked to the office (rent, insurance, office materials, etc.), the calculation was based on ERICs with

similar functioning. The strategy team did not forecast a major evolution between the construction and the operation phases.

2.4.1.1. Definition of roles

The work on the administration / coordination team started with defining the different roles needed to run an infrastructure like DiSSCo. In coordination with the CSO, a list of roles was made. It was based on the definition of the DiSSCo central hub, itself subdivided in different large components:

1. **Coordination and support:** support to DiSSCo governance and advisory bodies; financial coordination; human resources; monitoring and executing the SLAs and other contractual obligations with partners; coordination of the scientific and technical programmes; preparation of financial and work plan (5 years); coordination of strategic partnership (innovation and services development); general management; legal compliance
2. **Project development:** funding prospection, application, project management
3. **Community enhancement and training:** hosted by the nodes
4. **Facility coordination:** coordination of policies and processes
5. **Monitoring and quality control** (including standards policies programmes): performance / compliance
6. **Representation:** EU/international relations (incl. Expansion)
7. **Core service provisions:** coordination of provision of services, standards and policies
8. **User support:** immediate support at institutional level
9. **Communication and outreach:** public relations, communication

2.4.1.2. Definition of staff positions

All of these functions are linked to different staff positions. The assumption is that DiSSCo would need a minimum of 6 employees to fulfil these tasks:

1. **Director general (DG):** general management, preparation of financial and work plan, coordination of strategic partnerships (innovation and services development), general management, EU/international relations, Science-policy dialogue.
2. **Chief technical officer (CTO):** monitor and execute the service level agreements (SLAs) and other contractual obligations with partners, coordination of the scientific and technical programmes, coordination of IT, monitoring and quality control, coordination of services provision, standards and policies.
3. **Communication officer:** public relations and communication.
4. **Change manager:** coordination of policies and processes.
5. **Chief financial officer**: financial coordination, human resources, funding prospect, application, project management.
6. **Administrator, secretary:** support to DiSSCo governance and advisory bodies.

2.4.1.3. Wages calculation

A proposal is to calculate the wages of the team following European Commission salaries. The scales are available online and the figures represent a monthly cost including entitlements, taxation, social

security contributions as well as information regarding payment of the salary, including currencies and possible split of the salary. The financial entitlements are governed by EU staff regulations.[13] It represents the taxable gross pay per month.

According to the Netherlands regulations, the full economic cost of each employee is the multiplication of the taxable gross pay per month by 1.19 in addition to the taxable cost per month. This monthly full economic cost is then multiplied by the country correction coefficient (CCC) corresponding to the Netherlands. A coefficient exists for each European country and aims to correct the purchase parity differences among the different countries (see appendix 10). Finally, the annual cost per employee is the multiplication of the wage with CCC by thirteen (13 months).

X = Monthly wage from the EU

Y = CCC

Full annual economic cost per FTE

= ((X + (X*0.19))*Y)*13

2.4.1.4. Other costs calculation

Staff cost is a major cost for the central hub, but it is not the only one. According to the experience of other ERICs there is a certain number of additional annual costs to expect, such as:

- Furniture / office supplies
- Insurance
- Rent and charges
- Internet / mobile phones
- IT equipment for the team
- Other equipment / consumables
- Equipment for communication
- Production of brochures
- Organisation of events
- Advertising
- Bank account charges, credit card charges
- External auditor
- External legal consulting / advising
- Translation
- Website
- Training of DiSSCo personnel

[13] EUR-Lex - 52021XC1213(01) - EN: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.C_.2021.501.01.0011.01.ENG&toc=OJ%3AC%3A2021%3A501%3AFULL - accessed on November 14th 2022

These costs exist within each organisation. At this stage, there is no clear vision about the need for all these expenses. The figures that were calculated are guesstimates and could be different once the hub is in operation.

Here, it is also assumed that we are looking far into the future of the DiSSCo ERIC. The more we try to estimate needs for distant periods of time, the less easy it is for people to provide information. As WP4 does not have access to the current cost, it cannot be based on the current experience.

2.4.2. IT infrastructure cost calculation

The work on IT infrastructure and data preservation refers to milestone 4.7[14]. This document contains the first calculation of the cost to implement two main actions:

- Develop, elaborate and disseminate DiSSCo's core data model, the Digital Extended Specimen (DES)[15, 16], as a digital twin of a physical specimen in a NSC. A DES combines information from distributed and disparate sources about the physical specimen like sequence data stored in the International Nucleotide Sequence Database Collaboration (INDSC) databases or occurrence data stored in GBIF and forms in this way a new composite data type[17] for objects in a collection. Due to its design as an abstract autonomous unit of structured interlinked information, the DES is embedded in the wider framework of FAIR Digital Objects, which enables cross-domain interoperability of such units in data federation like European Open Science Cloud (EOSC)[18].

- Develop and maintain an innovative ecosystem of FAIR-compliant[19] e-services and tools for DES to facilitate production, mobilisation and deployment of specimen data (and metadata) from NSCs. The service network will provide essential key components of DiSSCo's Digital Specimen architecture (DS arch) including repositories to index and catalogue DES, services for community-based curation and enrichment of DES and infrastructures for minting and resolving of persistent identifiers (PIDs) for specimen data[20].

[14] Landel, S., Guiraud, M. (2022). IT infrastructure and data preservation indicators for DiSSCo RI Cost Book. DiSSCo Prepare WP4 – Ms4.7.

[15] Islam, S et al, (2020). Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections. *Data Science Journal*, 19(1), p.50. DOI: http://doi.org/10.5334/dsj-2020-050

[16] Alex R Hardisty, Elizabeth R Ellwood, Gil Nelson, Breda Zimkus, Jutta Buschbom, Wouter Addink, Richard K Rabeler, John Bates, Andrew Bentley, José A B Fortes, Sara Hansen, James A Macklin, Austin R Mast, Joseph T Miller, Anna K Monfils, Deborah L Paul, Elycia Wallis, Michael Webster, Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet, *BioScience*, 2022;, biac060, https://doi.org/10.1093/biosci/biac060

[17] Islam S, Weiland C, Addink W (2022) From data pipelines to FAIR data infrastructures: A vision for the new horizons of bio- and geodiversity data for scientific research. Research Ideas and Outcomes 8: e93816. https://doi.org/10.3897/rio.8.e93816

[18] Wittenburg P, Strawn G. Revolutions Take Time. Information. 2021; 12(11):472. https://doi.org/10.3390/info12110472

[19] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[20] Hardisty AR, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2021) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7: e67379. https://doi.org/10.3897/rio.7.e67379

The DiSSCo IT infrastructure – according to the diagrams shared by WP6 – is made of 23 components.[21] These components represent both services directly in contact with users (front office) and the "back office" that ensures the maintenance and smooth-running of IT services.

2.4.2.1. Perimeter: distinction between costs created because of the existence of the RI from the costs that would exist with or without the RI

This section is largely inspired by the methodology published by ESFRI in 2019[22]: guidelines on cost estimation of research infrastructures (see section on DiSSCo perimeter).

This indicator is important, notably regarding data preservation. At this stage, a working assumption is that DiSSCo will be a bridge to accessing the data preserved and produced by its member institutions. Such a scenario involves a cost to host DiSSCo software and components. Eventually, the time data is analysed and processed by DiSSCo components, it could transit in a server and would represent a certain number of terabytes but it would also represent immediate preservation. Such storage would not involve the storage of images and heavy data. It instead represents metadata related to these images. It is understood that this data will grow over the years: the more the RI will be strengthened, the more members it will have, the more collections will be digitised in Europe, the more DiSSCo IT infrastructure will have to process data and make them FAIR. According to that assumption, storage costs would grow over time.

Such a scenario means that with or without DiSSCo, data from institutions will exist and should be stored and preserved somewhere. With that reasoning, WP4 does not have to assess the costs of initial data preservation as, according to the counterfactual principle, it is not considered as a cost of the hub.

Still, as no decision has yet been taken on the matter, two scenarios are on the table:

1. Data storage is covered by DiSSCo Members;
2. DiSSCo could cover long-term storage in the case where institutions (notably small) do not have the equipment/budget to store - notably in the long term - their data. In that case, long term storage could be one of the costs of the RI.

When it comes to the DiSSCo perimeter and more specifically its IT infrastructure, teams should consider both back office and front office.

[21] Leeflang, Sam, Weiland, Claus, Grieb, Jonas, Dillen, Mathias, Islam, Sharif, Fichtmueller, David, Addink, Wouter, & Haston, Elspeth. (2022). DiSSCo Prepare D6.2 Implementation and construction plan of the DiSSCo core architecture (1.2). Zenodo. https://doi.org/10.5281/zenodo.6832200

[22] Centre for Industrial Studies (CSIL) (2019) Guidelines on cost estimation of research infrastructures European Strategy Forum on Research Infrastructures (ESFRI). Accessed on 2021-09-21.

2.4.2.2. Timeline: construction and operation phases

Figure 23 - DiSSCo RI timeline

To calculate the costs of DiSSCo ERIC, it is important to distinguish costs that will be invested to build the RI (construction phase) from the costs incurred to operate / run the RI (operation phase).

At this stage, the duration of the construction phase is estimated to be two years. Such information is not set in stone and from a cost calculation perspective, a proposed methodology is to calculate the costs to reach the Technical Readiness Level (TRL) 9 (maximum level - actual system proven in operation). At this stage, some DiSSCo e-services / IT infrastructure are being built as part of the DiSSCo Prepare and SYNTHESYS+ projects. They could reach levels 7, 8, 9 by the end of 2022. **The construction costs are the costs to go from the TRL reached at the end of EU projects to TRL 9.**

The calculation of the construction phase costs is based on the total number of cost units required to achieve TRL9. It means for instance that if you need two FTE(s) to build a service, you can either employ two people for a year, or one person during two years. The costs will be then adapted according to the funding DiSSCo will receive annually. The greater the funding, the faster it will be built and operational.

2.4.2.3. Staff cost: FTEs and average cost per FTE

From the first WP4 results, one observation is that most of the costs are staff costs. In order to calculate them, cost calculators have to know how many people will be employed, for how long and at which cost. This means:

Staff cost = Number of cost units * average price per cost unit

- Cost unit: full time equivalent (FTE) or PM (person month)
- Number of cost units needed: time required to develop and/or maintain the service
- Average price per cost unit (average full economic cost per employee)

The average price per cost unit could vary according to the type of job required.

2.4.2.4. Outsourcing cost: subcontractor / annual licences

From time to time, when the expertise / tools needed for a specific task are not within the organisation, it is possible to subcontract this or to buy a licence. This is part of the costs of the RI.

Outsourcing also represents cost units:

- Cost unit: cost per day, per package, per month, per year, etc.
- Number of cost units needed: time required to develop the service, packages required

- Average price per cost unit

Outsourcing cost = Number of cost units * average price per cost unit

2.4.2.5. Cloud computing / Hosting cost: storage (short and long-term)

As it could be expensive to invest in a data centre, RIs like DiSSCo can contract with companies to use their data centre to host their services. Two variables affect cloud services costing:

1. Expected computing power (measured in giga/tera/peta flops and used cores per hour);
2. storage space (measured in terabytes).

In that case, cost calculation can be done as follows:

Hosting cost = Number of cost units * average price per cost unit

- Cost units: computing power (number of giga/tera/peta flops) and storage space (terabytes)
- Number of cost units
- Average price per cost unit

Cloud computing relies on the type of service hosted. Some will require much more storage space and less computing power, and conversely, some will need more computing power than storage space.

With DiSSCo, three types of storage are identified:

1. Hot storage, one to store metadata from Natural Science Collections: potentially, as the IT infrastructure would process the data, it would need the capacity to store this data - at least during the time needed to process it. Such storage could represent some terabytes per year;
2. Hot storage: One to host the software and all components of the IT infrastructure;
3. Cold storage: One to store NSC data in the long term. This type of storage could be the responsibility of the member institutions. Still, according to the DiSSCo CSO technical team, long-term storage could be a service provided by DiSSCo, notably for small institutions.

Cold storage is less expensive than hot storage when it comes to storing the data but it is more expensive to access the data with cold storage than with hot storage.

2.4.2.6. Physical assets and depreciation

If large equipment, machines, etc. are bought for the research infrastructure, the full price is part of the construction phase costs, and renewal could be considered as a replacement cost[23]. Depreciation costs are not considered according to ESFRI methodology. Still, in the case where a large piece of equipment is bought, it is important to have information on its maintenance costs and an estimation of its duration: 5,10,20, … years.

[23] Replacement costs: correspond to the capital expenditure required to replace those assets whose economic lifetime is shorter than the reference period. The economic lifetime is different for various investment assets (buildings, machinery and equipment, etc.). In order to keep a facility in operation, each asset must be replaced at the appropriate time and the replacement costs, when foreseen since the start, must be included in the investment costs.

The purchase of a physical asset can lead to decommissioning costs: the cost to dismantle the asset has to be considered.

2.4.2.7. Distribution of the cost units among the different components of the IT architecture

According to deliverable 6.2, DiSSCo IT infrastructure will be composed of 23 different components, grouped in 9 containers. 'E-services' are part of this IT infrastructure as well as other components. This configuration is designed in order to make the infrastructure more sustainable and easily scalable over time.

At this stage, there is not a formal decision on the level of centralisation of this IT infrastructure. One working hypothesis is that some components, and notably e-services, could be developed and maintained by some of the RI members and not specifically the Central hub. This is how ELIXIR, a research infrastructure specialised in life science resources, is organised. Institutions (nodes) are located in different places in Europe and provide services on behalf of the RI.

In parallel, it is also explained that, at the beginning, DiSSCo ERIC may not have enough budget to build the whole infrastructure. The distinction between the different components would allow the ERIC to plan its budget and increase its investments over time: the greater the budget, the more it can go further in its developments.

The goal is then to estimate the distribution of the different cost units according to the different components of IT architecture. These results would help to distribute the investments required over time and, in the case where the services would be geographically distributed, among institutions.

2.4.2.8. User thresholds above which the components of the architecture should be redesigned (scaling of the infrastructure)

When it comes to IT software applications, there are user thresholds above which the whole system should be redesigned to cope with capacity. It could notably be the case for the Digital Specimen architecture and its capacity to support concurrent use of web services by many users at the same time. It means that the number of users has an impact on the cost to run the application and make it accessible because redesigning and scaling-up the IT system has a cost for the RI.

Here, it is relevant to consider the number of users that can be accommodated within the principles / indicators listed above. If the number of users exceeds this total, another type of system should be designed and developed.

2.4.2.9. Level of service provided by the IT infrastructure

According to the level of service the RI is engaged to provide, it could be more or less costly. For instance, if the RI declares that it will respond to any request in less than 24 hours, it is less expensive than if it declares a response in less than 2 hours. In the second option, it means that the RI would need to recruit more staff in order to have a larger team to quickly tackle all situations that could happen.

This rule should be mentioned in the service level agreements.

2.4.2.10. Training to engage users

In an extended understanding of the IT infrastructure, training is important to engage institutions / researchers to use the tools developed by DiSSCo and to share their data with the infrastructure. This cost can be either staff cost or outsourced (DiSSCo could prepare training materials and share it with subcontractors). It can be associated with change management and represent training hours but also large conferences during which the tools and their impact could be presented to all DiSSCo potential users. Deliverable 2.1, DiSSCo training strategy, explains the digital-oriented training activities for the future Research Infrastructure.[24]

These costs can sound marginal when it comes to IT development but such tools are often essential. A RI like DiSSCo is nothing without its users and members. It requires a budget to engage and train them on the tools the RI proposes.

Such costs could be decentralised: national nodes could train their members on DiSSCo IT tools.

2.4.2.11. Translation from operation costs to major upgrades

It is understood that major upgrades (depending on the service) updates are required at some regular time intervals and could represent a proportion of the costs connected with building the infrastructure (sometimes it is as expensive to upgrade a service as it is to build it).

At this stage, we could estimate that each year - as of the operation phase - 10% to 15% of the annual budget would be dedicated to upgrades. It means that even if upgrades do not take place each year, if there is a large one, DiSSCo would have reserves to cover it.

With this information, it is possible to make the following calculation:

Y = 10%

Z = annual costs - operation phase

major upgrades costs = **Y*Z**

[24] Sotiriou, S (et al.), 2022. Deliverable D2.1 "DiSSCo Training Strategy"

2.5 Limits of the DiSSCo Cost Book methodology

There are limits to this calculation and they are related to different variables:

- The concepts used can be misunderstood by the partners. There are explanations associated with the methodology but sometimes people do not read the whole guide and directly work on the table. It is even more of a risk when we do not communicate using the partners' native language. Some words can be misunderstood or misinterpreted. For instance, for some people, digitisation refers to the recording of data, sometimes without any pictures. Sometimes, it includes a picture. WP4 tried to mitigate against this risk of misunderstanding by having bilateral meetings with participating institutions, upon request.

- The accounting entities are not always coherently subdivided. For instance, several institutions used one table for their entire collections department. It means that no difference is made between insects and vertebrates for digitisation, for example. This can lead to bias and provide costs that do not fit with reality.

- The costing methodology can be time consuming and people can enter inaccurate data in a rush to finish the activity.
- The use of Excel tables for the DiSSCo costing methodology can sometimes create problems. When the data is copied and pasted, the formulae can disappear. There is a bias when you share a tool and the way people can interpret it. A possibility is to harmonize the training in the used software amongst the users.
- As explained above, depreciation is not accounted for and, according to the reference year (2019), the investment within an accounting entity can be really high. This can have a major impact on final calculation and increases the annual costs. This would result in inaccurate data.

- Even now, at the end of DiSSCo Prepare, the DiSSCo level of service is not completely defined. This means that some strategic planning has still to be refined in order to develop the DiSSCo implementation roadmap. Without this information and a clear plan, the cost calculation may become irrelevant over time, notably regarding the IT infrastructure and the central office costs.

CHAPTER 3. Results

Task 4.1, dedicated to DiSSCo Cost Book development, has used the methodology described above to calculate the costs of the DiSSCo Central Hub Office and IT systems. A first estimation is that the construction phase would cost around €M 2.5 per year if it lasts two years. DiSSCo operation would annually cost €M 2.6. At this stage, some of these costs can be in-kind or cash from member institutions. The costs to decommission the RI would be around €M 2 to 4 (1 to 2 years of operation).

Meanwhile, due to the fact that WP4 is working on costs of a RI which does not exist yet, costs are significantly distributed to the decentralised implementation plan of the RI. In that context, the results from this present DiSSCo Cost Book are partial. 27 institutions out of 170 responded to the methodology shared. It is understood that even if the team tried to ease the process as much as possible, it is still a costly and time-consuming exercise and not all contact persons had time and interest to implement the Cost Book methodology. Another analysis is that the so-called "national nodes" are not all yet mature to share and gather results. They were designated as contact points in order to distribute the methodology more broadly. Still, most of them did not respond and only a few distributed the documents with their networks.

3.1 DiSSCo Central Hub Office costs

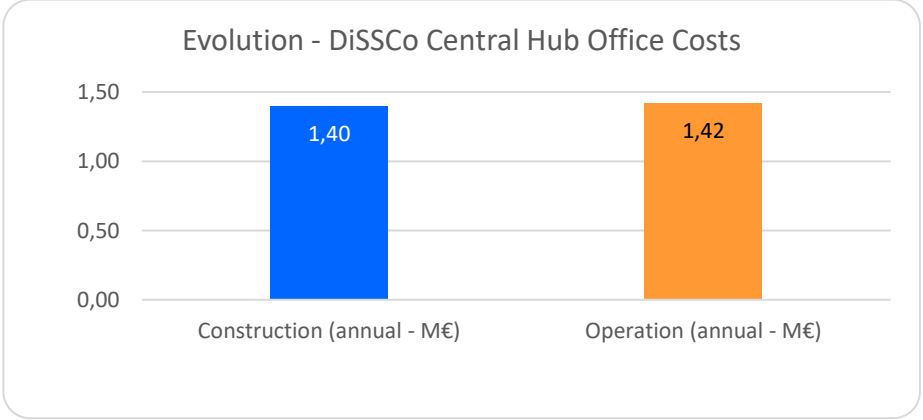
The Central Hub costing raises different questions: the evolution of its activities between the construction and operation phases, the nature of its functions, team composition and the nature of its costs. For DiSSCo RI, the assumption is that it will be mainly a team, located in one place, who will coordinate the implementation of the RI.

3.1.1. Evolution between construction and operation phases

According to the first estimation, the costs to coordinate/administrate the research infrastructure will not vary significantly between the construction and the operation phases. This understanding is applicable while the RI receives the same annual funding.

Once DiSSCo is in operation, with a dedicated team, annual funding and formal statutes, it can implement an efficient business model to increase its activities. Here the costs reflect the minimum budget the RI needs in order to be operational.

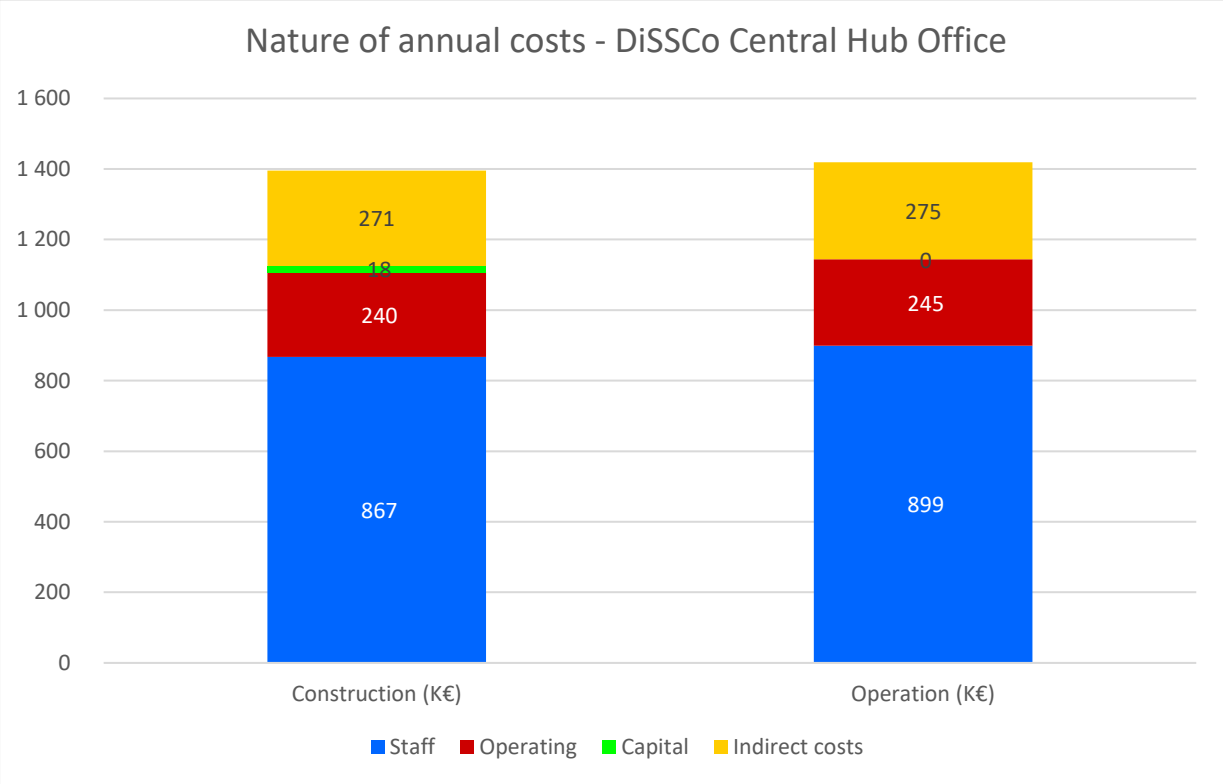
Figure 24 – DiSSCo Central Hub Office - Cost evolution



3.1.2. Nature of costs: mainly staff costs

The costs of the DiSSCo Central Hub Office will be mainly staff costs (62%). Some operating costs are also calculated and have a lower importance in the total budget (17%). Capital costs are low as DiSSCo does not require the construction of a physical infrastructure; it is about providing digital access and facilitating physical access to natural science collections held all over Europe. Such an assumption can evolve in the scenario where DiSSCo initiates the physical construction of centres of excellence. Its implementation therefore depends heavily on staff who will be mobilised to both coordinate the stakeholders of the RI and design the IT tools that will facilitate access to the institutions' data. Finally, indirect costs are calculated based on the EU rules: 25% of the direct costs (excludes subcontracting).

Figure 25 - DiSSCo Central Hub Office - Nature of costs



3.2 DiSSCo Central Hub, IT infrastructure and e-services costs

3.2.1. Evolution between the construction and the operation phases: increased investments and lower running costs

As mentioned within section 2.4.2.2, “At this stage, the duration of the construction phase is estimated to be two years. Such information is not set in stone and from a cost calculation perspective, a proposed methodology is to calculate the costs to reach the Technical Readiness Level (TRL) 9 (maximum level - actual system proven in operation).” In that sense the total costs to build DiSSCo IT services and architecture is estimated at €M 2.2. It means that if the construction phase lasts two years, each of them would cost €M 1.1 (see figure 26).

The cost to annually run and maintain the DiSSCo IT services and architecture is estimated at €M 1.2. This is an annual cost and the duration of the construction phase has no impact on this figure.

Figure 26 - Evolution of Total DiSSCo Central Hub IT costs between construction and operation phases, ***based on the hypothesis that the construction phase will last 2 years***

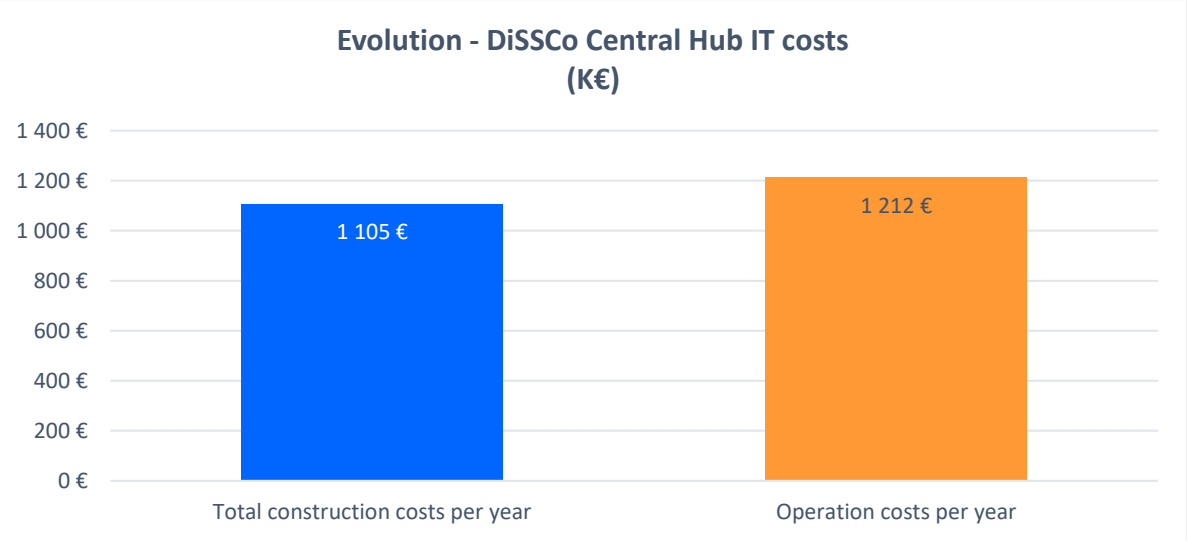
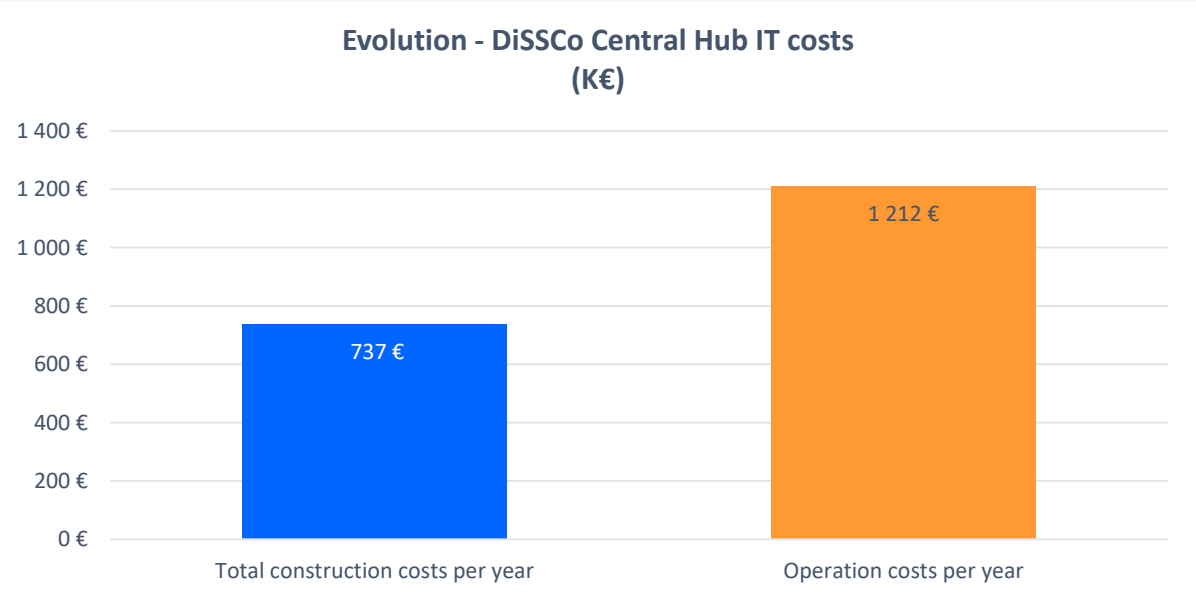


Figure 27 - Figure 6 – Evolution of Total DiSSCo Central Hub IT costs between construction and operation phases, ***based on the hypothesis that the construction phase will last 3 years***

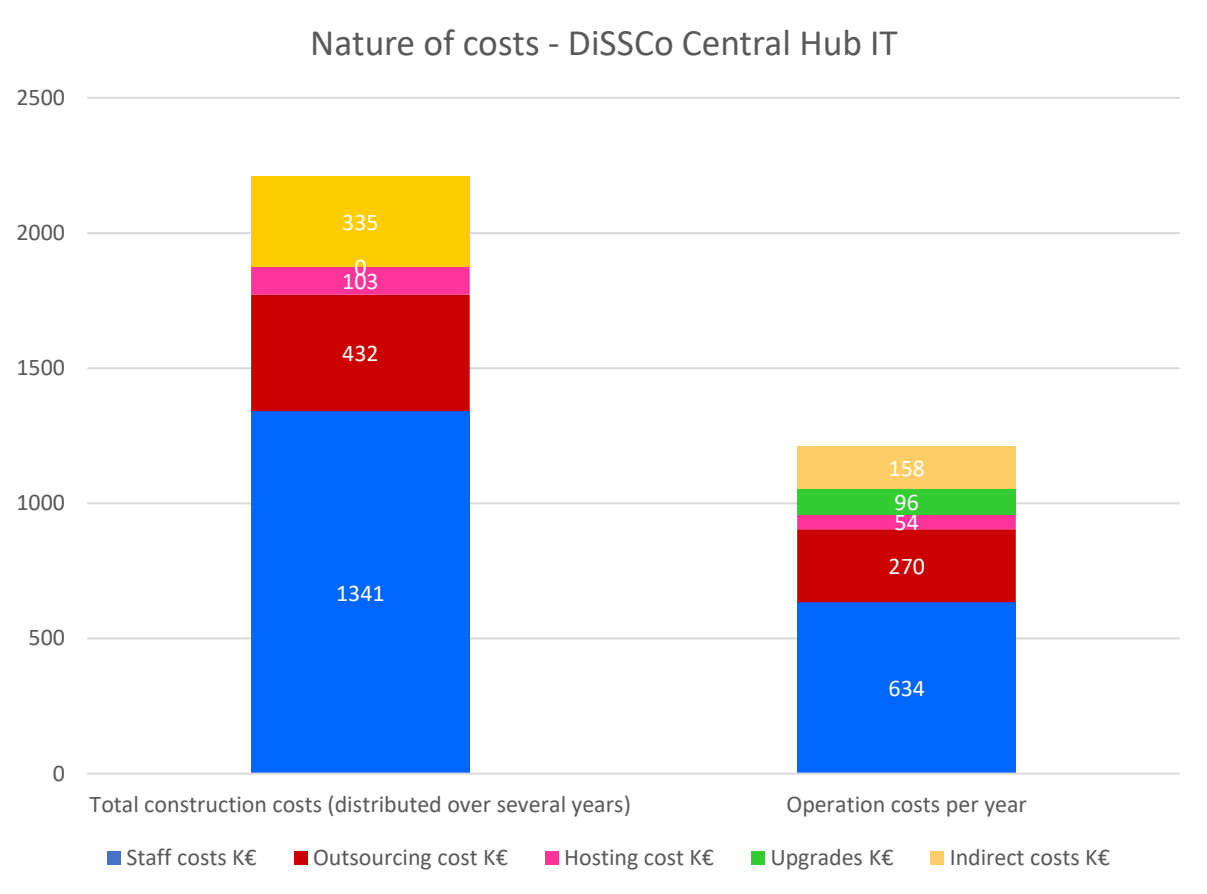


3.2.2. Nature of the costs: mainly staff costs

The nature of the costs can reflect the way the work is organised and implemented. Regarding the Central Hub IT infrastructure and the development of e-services, most of the activities are planned to be implemented in-house. Most of the cost calculated are staff who will be employed by the DiSSCo hub – including by institutions who have signed SLAs with DiSSCo ERIC. Under both phases, DiSSCo

could delegate some of its tasks to subcontractors for some occasional and specific work. The cost section dedicated to subcontracting encompasses costs related to membership fees such as ORCID and DOIF. Costs can also include annual licenses for specific software, although the plan is to mostly use open source software. Hosting costs are relevant in the case data storage would be outsources. In that scenario, the more DiSSCo would be responsible to store NSC data, the more its costs would rise. Finally, upgrade cost is estimated to be 10% of annual cost (Section 2.4.2.11).

Figure 28 - Evolution, nature of costs - DiSSCo Central Hub IT, based on the working hypothesis that the construction phase will last 2 years



Within deliverable 5.5[25], "Construction plans for the improvement of technical infrastructure in the areas of geo-collection data and taxonomic", there are some details about related to some of the DiSSCo RI IT infrastructure.

[25] Woodburn M. et al. (2022): DiSSCo Prepare Deliverable D5.5 Construction plans for the improvement of technical infrastructure in the areas of geo-collection data and taxonomic services. DiSSCo Prepare. Page 37. https://doi.org/10.34960/dzs0-xa94

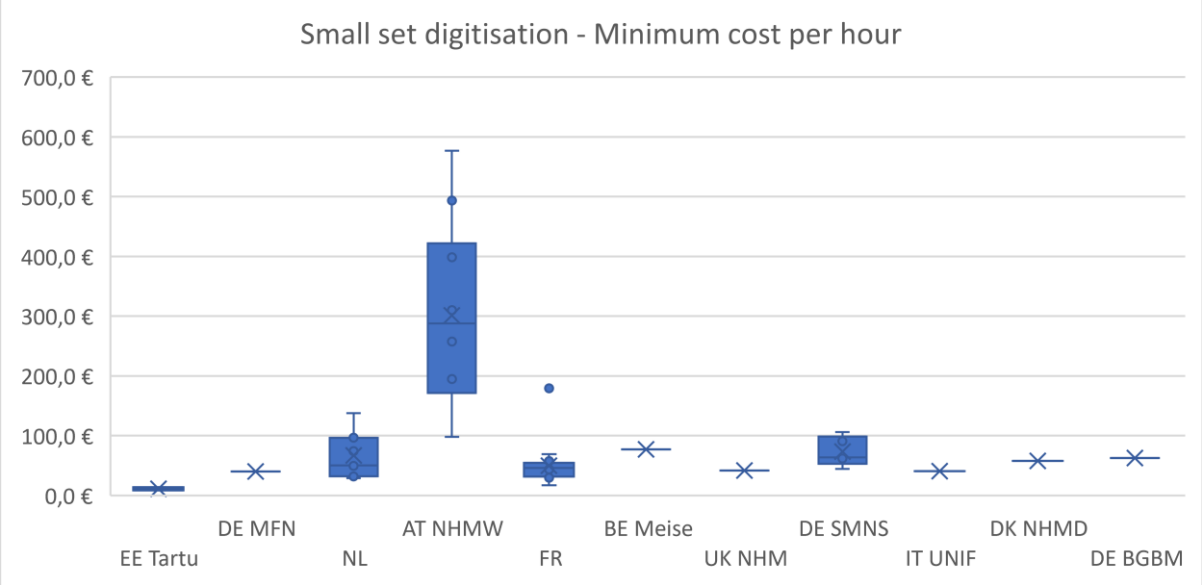
3.4 DiSSCo national nodes, cost per hour

3.4.1 Limits to the cost methodology

27 institutions from 9 countries (AT, BE, DE, DK, EE, FR, IT, NL, UK) have answered the Cost Book survey (including MNHN results) totalling information on 76 accounting entities.[26] Information is missing for some parameters but on the whole this survey constitutes a good basis for an overview of the costs within the RI.

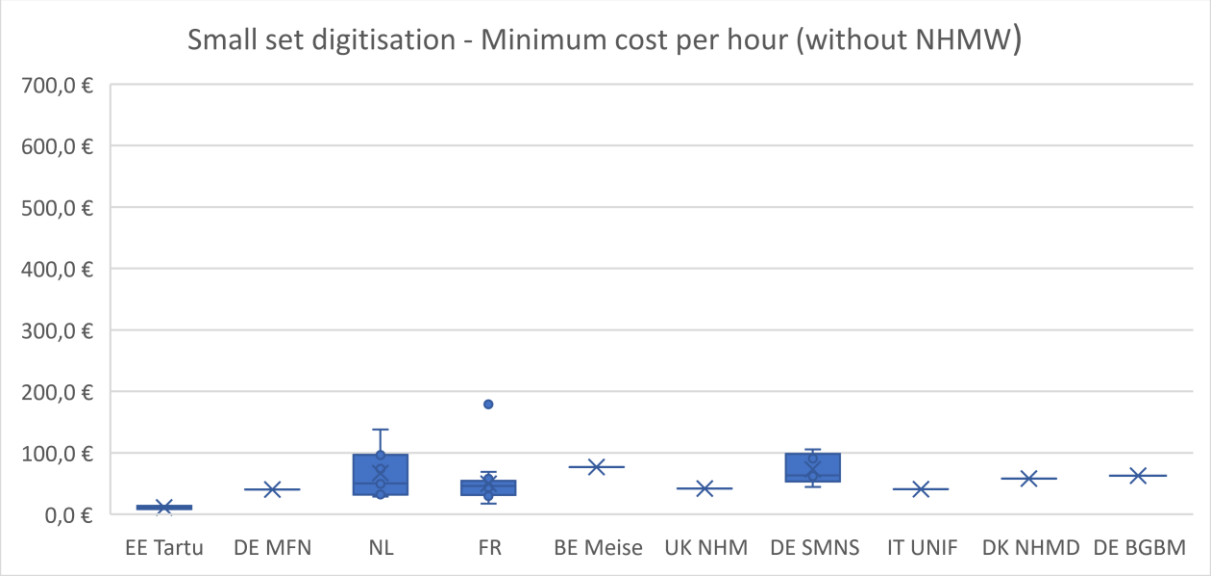
The first conclusion is that there are discrepancies whatever the parameter taken into account. Two examples are shown below (Figures 23/24 & 25/26).

Figure 23 - Minimum cost per hour for small set digitization



This figure shows that NHMW data are significantly higher than all the others.

Figure 24 - Minimum cost per hour for small set digitization after removing NHMW



[26] See Appendix 11

Once the NHMW data are removed from the dataset, the values are more akin to a normal distribution. However, the spread of the data remains quite large, with minimum-maximum values ranging € 10-100 if two more outliers are removed from the dataset.

Figure 25 - Minimum cost per loan

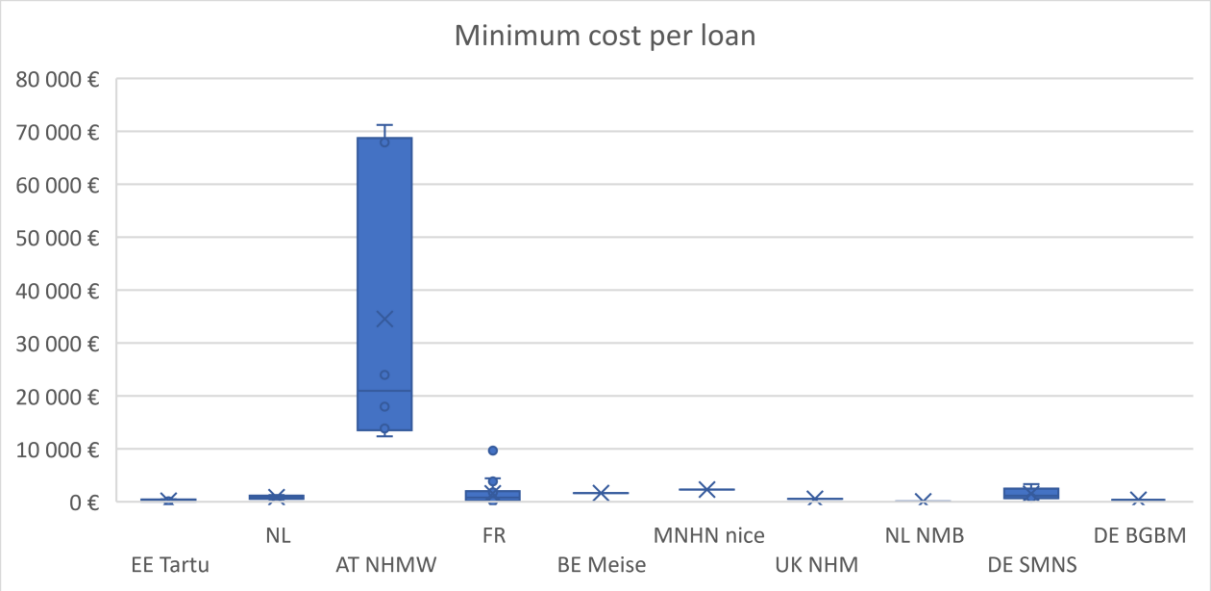
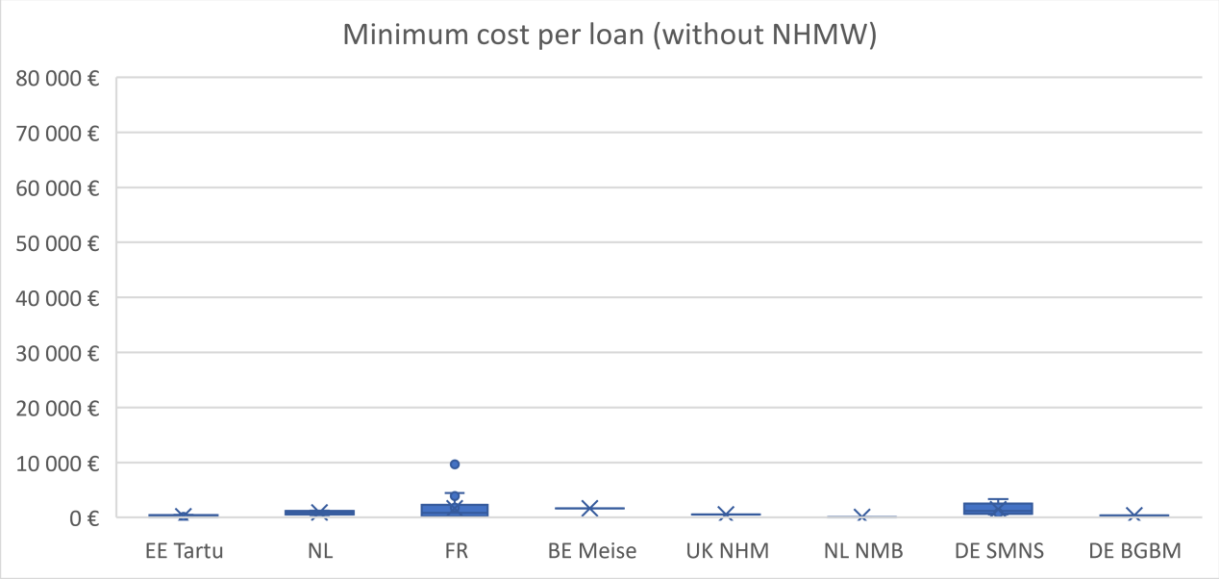


Figure 26 - Minimum cost per loan after removing NHMW

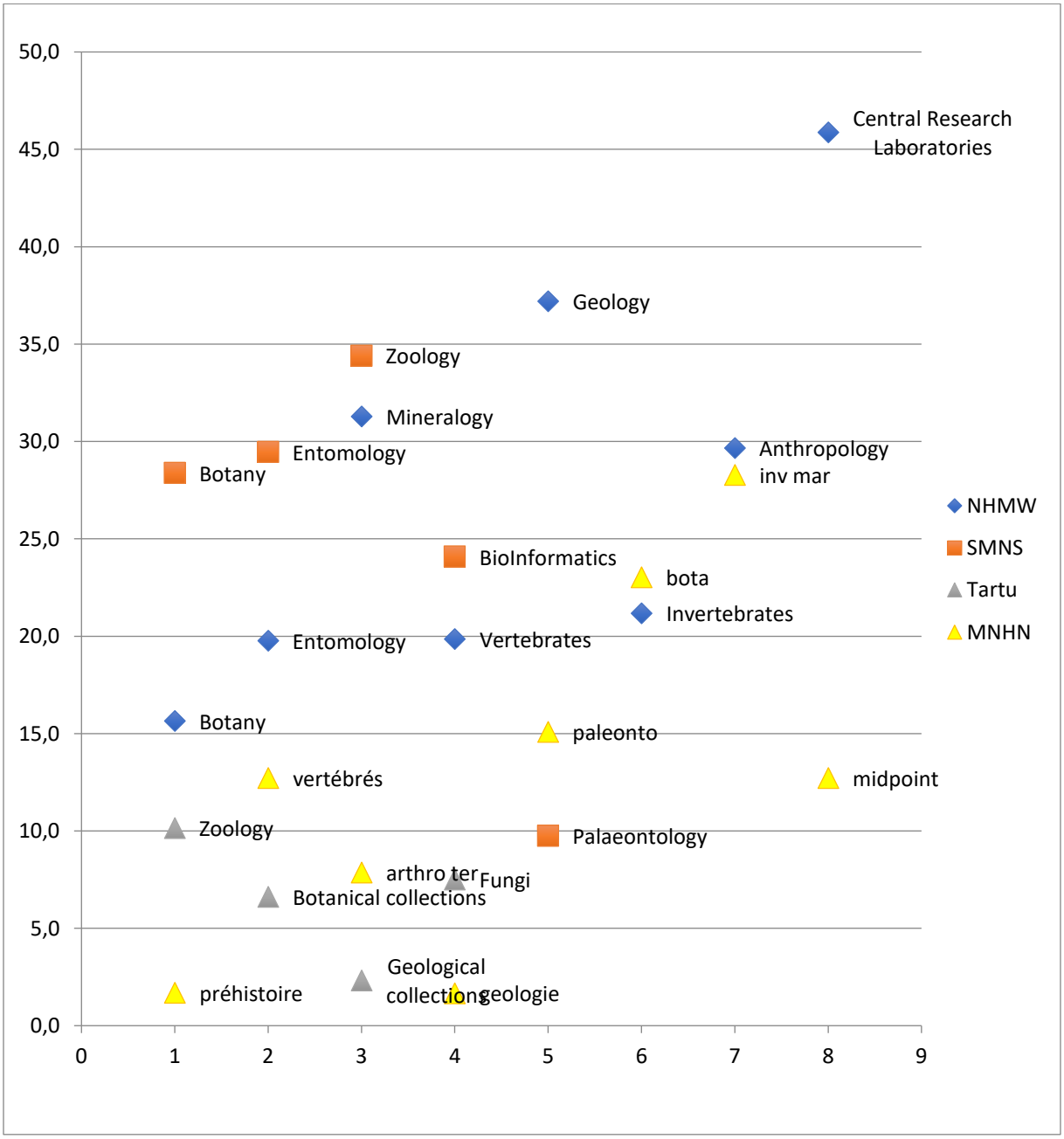


As for the cost per hour for small digitization, NHMW data stand out as outliers, but removing them from the dataset still leaves a significant spread of data.

Further work is needed to check whether this discrepancy originates in structural differences reflecting the true cost differences or are errors/mishandling of the institution's data in filling the Cost Book. As a consequence, only median values were used as it was found that they account better for outliers in a series than trimmed means (unless removing both outer quartiles in the present case).

Using hours as a unit to measure services alleviates the problem of differences due to cost. Furthermore, it allows comparing different AE's within an institution without the bias between institutions shown above.

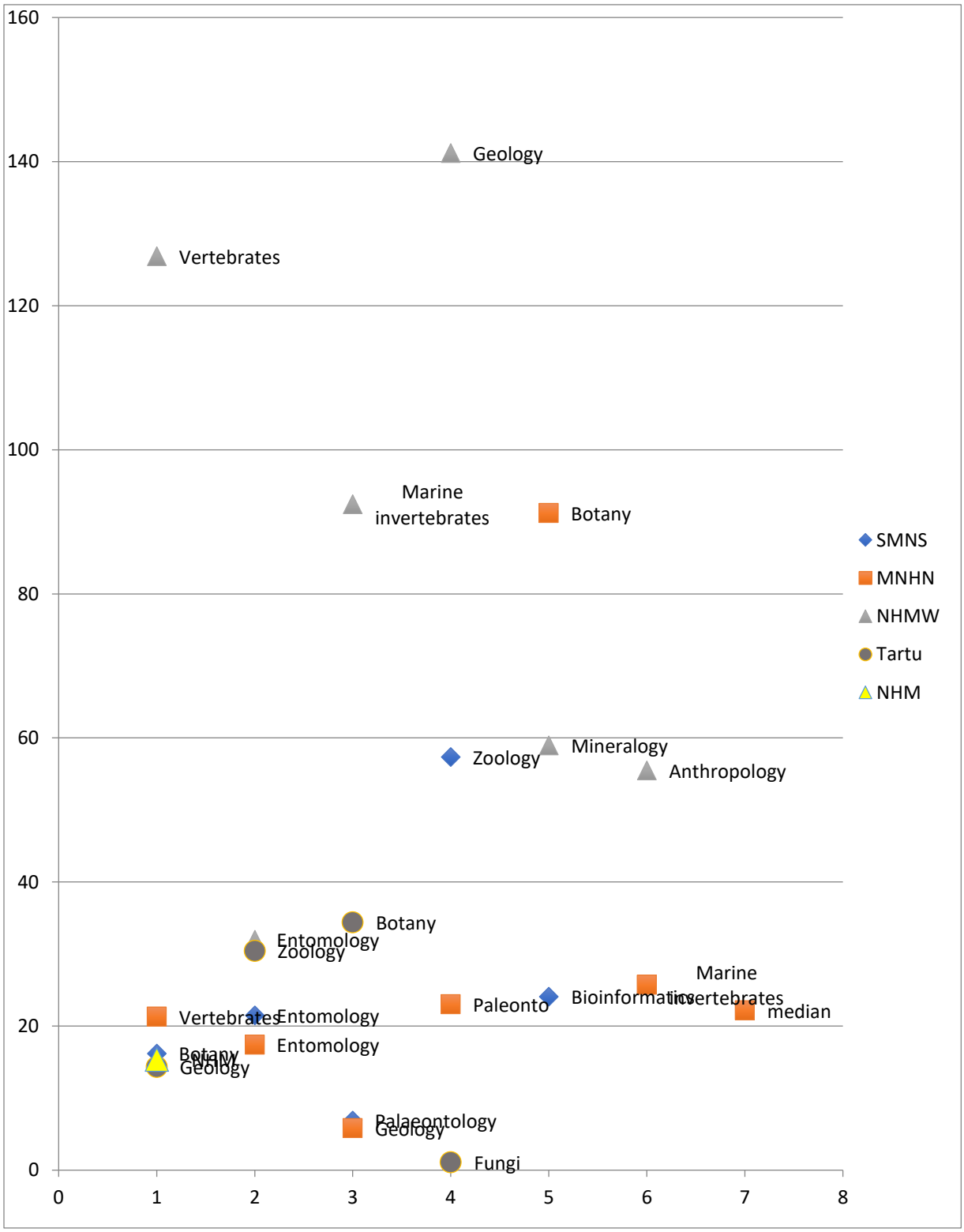
Figure 27 – Number of hours per visitors for each Accounting Entity (AE) in the institution



This figure shows that the data are more evenly distributed; although NHMW is still on the high part of the diagram there is no clear evidence that there is a structural element in the high cost reported by NHMW as they match SMNS values.

Another element shown by this diagram is that there is no pattern in AE's within each institution: for example, at MNHN, geological AE reports the lowest number of hours and botany reports the highest, whereas it is the contrary for NHMW. The spread of the data within an institution is not significantly different from the spread observed among all the institutions.

Figure 28 – Number of hours per loan for each AE in the institution



The figure for the loans (Figure 28) shows the same pattern as hours per visitor (Figure 27). As for the cost data (Figures 25/26), the median is the best indicator of the distribution of the data. It should be pointed out that the median for MNHN and the value for the whole collection management unit of NHM UK are about the same, which is expected for two institutions of similar sizes.

3.4.2 Main results

The proportion of staff costs in relation to total costs is 90%. The differences between individual data cannot be ascribed to differences in the nature of AE. Understaffing of some collection departments and high share of equipment cost in analysis unit would have the same effect for lowering that percentage.

Comparing the median cost of the services is consistent with the assumption that the structure and nature of staff are different with respect to the service.

Table 1 - Median values for the different services of all responding institutions. Values in brackets correspond to the median after removing NHMW data.

€/hour	Mass digitization	Small set digitization	Consulting services	Training	Analysis	Loans	Physical access
Minimum	67 (48)	61 (51)	64 (57)	68 (65)	65 (65)	57 (48)	59 (54)
Maximum	81 (61)	74 (64)	84 (70)	87 (79)	82 (82)	69 (63)	73 (67)

Consulting services, training and analysis generally require more seniority and specialist skills than digitization, loans and physical access, a difference which will be weighed on the cost per hour of the staff. The reason for mass digitization to be the same as consulting, training and analysis is due to a smaller set of data (only 14 institutions of the 17 responding institutions reported mass digitization costs) and the important weight of NHMW in this small dataset which represents 10/16 data. The values in brackets which correspond to the values after removing NHMW data are more consistent with the above assumption.

It is important to have central values for the cost of services as a practical tool to assess the feasibility and the cost of users' projects. The values calculated from the survey can be compared with the data derived from the pilot study at MNHN. The table below shows that there is a good correspondence between the both surveys.

Table 2 - Comparison of median values for the different services between MNHN and the other institutions.

	Unit	All institutions	MNHN
Mass digitization	Min-max cost per specimen	2.92 - 3.40 €	
Small set digitization	Min-max cost per specimen	12.68 - 14.86 €	12.25 - 19.84 €
Loan	Number of hours per loan	22 hours	22 hours
	Min-max cost per loan	1107€ - 1306€	1379€ -1800€
Physical access	Number of hours per visitor	17 hours	13 hours
	Min-max cost per visitor	1068€ - 1402€	873€ - 1203€
	Min-max cost per visitor day	312€ - 409€	241€ - 299€

As practical tools, it is worth noting that mass digitization per specimen is 4 times cheaper than small set digitization. The values obtained for the cost of a visitor per day are of the same order of magnitude as the costs in the Synthesys+ program, a circumstantial evidence that the methodology proposed here is consistent with other methods of cost assessment within the RI.

3.4.3 Conclusions on the results of the consultation survey

The distribution of the data within and between institutions is large. There is no unique cost for the RI and therefore there cannot be a unique fee to users across DiSSCo.

The data support the assumption that most costs are staff costs. They represent about 86% of the total cost. Therefore, it could be possible to approach the cost of the services by knowing the distribution of staff time and the full cost of staff to the employer. However, hourly costs need to be adjusted for AE's and services in order to reflect the granularity of the institution costs.

Median values for key services have been estimated and can be used as first approximation to projects.

Further study is required to analyse processes behind the services in order to eliminate the bias created by the procedures (digitization associated with loans, different MIDS level of digitization, etc.)

3.5 Total cost of European Natural Science collections

One aim of the Cost Book exercise was to estimate the cost of the DiSSCo RI as a distributed infrastructure. Estimating this cost will bring some light on the cost generated by creating the central hub. This estimation is partial as it is based on data from only 27 institutions[27] (out of 170). Besides, as shown above, the diversity of the institutions and the way they understand the Cost Book makes the extrapolation to the 170 institutions questionable. However, we can take a conservative approach to estimate a minimum cost of the entire DiSSCO RI by not including the costs reported by NHMW (see section 3.4.1). With this due caution, the services provided by the institution's members of DiSSCo would be around €M 290 per year (including the cost of maintaining the collections). Whatever the uncertainty of this value, it clearly shows that the full cost of the hub during the construction phase and the operation phase represents only around 0.9% of the total cost needed for building and running DiSSCo RI.

Table 3 – Calculation of the DiSSCo costs declared

	Total annual costs declared (without NHMW) €
1. Mass digitisation	4 713 293
2. Small set digitsation	3 302 777
3.Consulting services	2 866 473
4. Training	1 192 179
5. Analysis	9 213 017
6. Loans	2 176 605
7. Physical access	2 909 885
8. Asset maintenance	17 743 182
Total DiSSCo costs declared (sum 1-8)	44 117 412
Total costs declared	73 336 429
Total non DiSSCo costs	29 219 018

DiSSCo RI estimated annual costs calculation

$C = D*(T/R)$

$C = 44\ 117\ 412*(170/26)$

$C = 288\ 459\ 999$

Whereas:

C: DiSSCo RI estimated annual costs

D: Total DiSSCo costs declared by participants

T: Number of institutions in the DiSSCo perimeter

R: Number of institutions who participated in the WP4 Cost Book

[27] Details within Appendix 11

3.6 Detailed study on training costs

3.6.1 Introduction of the study

Under the DiSSCo Research Infrastructure overarching umbrella, training stands as a building block to enable agents involved in its creation and development, as well as the different potential user communities, to access and utilize the Research Infrastructure to its full potential.

The combination of these two streams for capacity generation and improvement illustrates the variety and complexity of the exercise to identify a unit cost for training services.

The scope of the definition of Training Cost is strongly related to the DiSSCo Training Strategy (Deliverable D2.1) that aims to present how digital transformation within an organization leads to training at all levels, across competences, and over time. This consequence of phases that drives development of the digitization work from an early, focused and isolated task, to a broad, comprehensive, and multidisciplinary role that also evolves over time, thus adding proficiency through learning and experience to reach high level management competencies. The Strategy stands as a step further from the Conceptual Design Blueprint developed under the ICEDIG Project [28] which claimed for a more intensive, multidisciplinary, coordinated, and periodic training to tool-up staff in natural science-collections-holding institutions, to be able to successfully intervene in and make efficient use of the DiSSCo Research Infrastructure at its full potential.

The currently presented Cost book aims to respond to a particular cost line to be introduced in the final DiSSCo Cost Book, to frame the initial budget that the Research Infrastructure needs to cover to be fully operational, while being able to provide its core services, at its minimum stage of development.

When tackling training and in the attempt to analyse how much would training cost in the DiSSCo landscape, a multitude of parameters need to be considered, each of which will have a different degree of influence and direct effect on the composition of the training cost.

Under T4.1, the objective of identifying the cost of training (being considered a pivotal service for DiSSCo) has evolved over time as the number and diversity of features enlarged and the relation among them become more and more interlinked.

3.6.2 Training Cost

By *Training* we refer to all “formal” (stable, planned and structured) learning actions (“courses”), provided by all type of trainers, in all formats, from external or internal resources, which objective is to provide, enhance and enlarge skills of the staff members of the institutions involved in the CETAF/DiSSCo community. This training will allow for enabling them to make effective use of the RI, as both providers and users of the information and services. We do not consider any possible induction activities e.g. planned time to receive introductions to aspects of the work.

[28] Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalga A, Paul DL, 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: e54280. https://doi.org/10.3897/rio.6.e54280

Identifying the unit has been equally a challenge since either of the approaches imply a large set of unknowns. From the concept of a 'course', to the number of either trainees or trainers, there are many variables and uncertainties that impede progress towards harmonizing and correlating unit costs in an unified manner.

As a premise, this work started with the course as the unit element from which to analyse the main components in financial terms. The unit course was then abandoned and the time unit in days was selected. It easily expands or reduces the time coverage of the course (by multiplying or dividing the basic unit cost). The unit "day" conforms better to the different sub-costs assigned to training in person, where the highest support and indirect costs apply. Another principle at the basis of this current study is that training runs at its full capacity both in time (full days) and coverage (full attendance). Doubtless, the latter premise considerably influences the income received, should the course have a fee. However, the present analysis is solely focused on the cost side without considering the level of income or potential profit gained with the training actions implemented.

From all the above principles and premises the first attempt to tackle how to identify "the" training cost started with the breaking-down of the final cost into its multiple components.

3.6.3 Methodology

1. Matrix Components

Initially the work was planned as to identify the Cost Components by developing a Matrix that could encompass all major parameters to be considered including:

a) Typology of costs:

- Direct
    - Personnel
    - Operation
        - Facilities
        - Catering
        - Support costs (travel and accommodation)
    - Documentation
    - Communication
        - Advertising
        - Registration
- Indirect:
    - Governance, steering and management
    - Infrastructure
        - Digital system
        - Venue / real estate
    - Others

b) Category of the unit:

- Content
    - Create (start developing skills)

    - Upgrade (gaining proficiency and enlarging competences)
    - Multidisciplinary (across roles and workflows)
- Length
    - Duration (of each course /unit)
    - Recurrence
        - Punctual
        - Periodic
- Format
    - Environment (on which the course is allocated)
        - Online
            - Streamlined
            - Open
        - Physical (in presence)
        - Blended (combined physical and virtual training)
    - Dimension
        - Plenary (addressing a generic audience)
        - Dedicated (specialized, on specific topics and/or areas)
- Providers (Trainers categories)
    - Area (5 Dimensions) since different areas may require distinct expertise
        - Technological (in tools, processes, mechanisms, workflows, and all means that support digitization efforts)
        - Organizational (for a better and comprehensive understanding of the organization around digitization, including recognition of applicable legislation, monitoring and reporting, work structuring and engagement with other sections, both inside and outside the organisation itself, as well as career development and other related HR issues)
        - Financial (to add complementary competences in areas such administration, compilation and costing of data management and/or digitization work)
        - Data (and all topics around data from its generation to publication, use of standards, models and protocols, data architecture and databasing, data curation and management and storage, and data quality, which all together will allow harmonization of work inside and outside the institution)
        - Scientific (to support and improve the research run around natural science collections and promote the linking paths between the physical and the digital objects).
    - Seniority (professors, lectures, associate professors, PhD students, other categories from the national ranking systems)
    - Affiliation (in relation to the institution)
        - Internal
        - External
    - Type (researchers, technicians, other categories and sub-categories)
- Audiences (Trainees composition)
    - Type (students, professionals, museum staff)
    - Scope
        - Strategy (broad landscape)

        - Specialized (in areas, objects and/or steps or methodologies)
        - Operation (in the implementation phases, from preparation to handling and imaging)
        - Support (from communication to laboratories)
        - General (around digitization)
- Material
    - Digital material (available online)
    - Physical documents (printed copies)
    - Helpdesk (in support on either of the above)
- Awards
    - Certificate
    - Credits (academia)

Moreover, most of those categories could be split into “own” (resources or personnel) and “procured”, which adds another layer of complexity.

Combination of all those variables tended to be unfeasible to be meaningful and to provide a fair response to the key question: how much would be training provision cost?

2. Additional variables: Diversity across countries

To the complexity detected in formulating a reasonable matrix that could be applicable homogeneously across DiSSCo participants, the research brought in an additional variable that should be considered towards a further analysis.

Provided that the major component of the final Cost analysed under the categories listed above came from the Personnel involved in the provision of the training, salaries of the training providers became the critical cost line around which the cost could largely fluctuate from one country to another (and even at a more granular level, within the same country).

Therefore, the analysis would have to integrate an indexation by country. Moreover, that indexation should strongly refer to a certain moment unless another index is added as per the time frame considered for the training provision. Current inflation rates across Europe (and beyond) clearly shows the disparity of costs that could emerge if the analysis is made in the beginning of 2019 or in late 2022. Unforeseen but highly influential events could have a strong effect on the final calculations.

On top of this inflation effect, the multi-lingual approach, if considered, would integrate another facet that would lead to additional costs depending on the country where the course is provided (participants-wise). For our analysis, we have assumed that all training is provided and accessed in English, based on the facts that:

- English is the language regularly used in the scientific community
- It constitutes a vehicle for harmonization across DiSSCo community
- It allows expansion and ensures broad access.

Despite those arguments, some institutions, when supporting the learning process and career paths of their staff, may need to translate the material into their national languages but the implied costs for such action is not considered here.

3. Change in perspective

Specific research was made and analysis of different institutions based on the data provided were carried out. The resulting outcomes were critically uncertain, confusing and ambiguous in many aspects:

- Consistency: results obtained from different institutions in different countries are too inconsistent to obtain a fair average for any training unit.
- Comprehensiveness: disparity among the large diversity of training courses also prevent us from reaching a consistent formula to determine the cost of training as a unique figure that could serve one for all.
- Lack of uniformity: the different dimensions to be tackled, as well as the levels of expertise expected to gain, multiply the differences between the training units.
- Unevenness: cost differences across borders lead to a huge discrepancy among the results obtained, even with the national-based indexes applied.

From all the above, it was clear that the approach to determine a cost line that could be aggregated and integrated into the final Cost Book for DiSSCo RI was more than a challenge, a target difficult to express in a single (or a set of) figures.

A new approach was required that could allow building a thorough, well-based, and harmonized model.

It was decided to change perspective entirely. Instead of looking at the ending unit as the cost that training implies from the provider side, the new approach adopted aims to establish the cost in terms of how much the training user is willing to pay to receive the training unit.

This perspective dilutes most of the difficulties by considering stable and publicly available data unified at European level. Variables such as country differences, timeframe, trainers' type and categories disappear from the analysis.

The basics for this new approach lay on the ERASMUS+ Programme run by the European Commission for 35 years.

This framework has proven as a very reliable and meaningful programme for education and training. In the ERASMUS+ Annual Report 2020 (last available) the following numbers give idea of the importance, extension and impact of this programme dedicated to training:

Budget: 3.8 Billion €

Participation:
- nearly 640.000 people trained, studied and volunteered
- 126.900 organizations involved

Implementation: through more than 20.400 projects

4. Opportunity cost approach

The study therefore changed perspective and we moved from the side of purely counting the expenses of providing a training unit (provider cost) to the fact of how much is being currently paid for granting

access (trainee grants). **As per the figures compiled from ERASMUS+, the programme pays 80 EUR per participant per day.**

The average number of 600 trainees per year is obtained from the following statement:

We assume that 20-21 courses are organized per year considering a number of three per country of the participating organizations in Task T2.1 involved in the definition of the DiSSCo Training Strategy (i.e. Greece, Norway, Belgium, UK, Portugal, Italy and France). Those courses will allow staff from all countries to participate in the training actions reaching out to 30 participants per course.

On top of this basic figure of reference, the model continues with several estimates that aim to provide harmonization and to unify disparity of environments and situations. To that end, once DiSSCo becomes operational the Training Service will be implemented accordingly. Still, to allow the large community of DiSSCo access to the varied offer and to enable the Training Provider to ascertain priorities (of subjects and audiences), Training is considered to be implemented in three (3) subsequent phases over the first 3 years of DiSSCo Operation. Each of those phases include an estimate of the number of trainees and the number of courses offered. **For the purposes of the current exercise, an average of 600 participants per year is considered.**

3.6.4 Training Implementation

1st phase (year 1): The estimation is that in the first phase of implementation about 200 participants will participate in about 10 courses that will be organized by DiSSCo (through the centre of Excellence credited to be DiSSCo Training Provider).

2nd Phase (year 2): the number of trainees is expected to be increased to more than 500 attending a number of about 20 courses.

3rd Phase (year 3): the forecast is to engage with 1.000 trainees who will participate in 50 training courses provided by the DiSSCo Training Provider per year.

The feedback obtained during this timeframe of 3 years will substantiate the development of a valid working plan for training activities.

Training Service Provider: although selection of the final Training Provider for DiSSCo Research Infrastructure has yet to take place, this model is based on the long experience of the DiSSCo Prepare partners in the organization of such training courses and fundamentally, in the operation and framework of CETAF-DEST, which safeguards the validity of the proposed training plan.

Training Cost: as per the ERASMUS+ Programme (Action KA1), the cost of the course per person per day is set-up as €80. Provided that an estimate for the course length is set at 6 days, the cost per trainee amounts up to € 480. This amount shall cover all expenses, both direct and indirect, of the course.

On the basis of an average number of 600 participants per year, the final cost for Training staff in DiSSCo will be €K 290. This number of trainees include all levels of staff, from managers to PhD students.

On top of this calculation, other indexes and considerations may need to apply:

- **Inflation rate**: the average rate in Europe is currently 10.0% in the Euro zone[29]. Therefore, the cost line presented here for training should be increased by this inflation rate to a final amount of €K 319 per year.

- **Online format**: this can provide several advantages over physical attendance, such as higher flexibility, large range for adaptability, continuous learning, recurrent and quick upgrade, easier trainers engagement and others. Still, in the field of digitization, in-situ training offers the highly valuable experience of performing activities with real objects, understanding processes and procedures in real time, and having mentoring from the practitioners and many others.

Based on several case studies provided by NHM London and MNHN Paris, and assuming the cost and dedication of the trainers remain same (for the preparation, organization and training) despite the format of the course (either physical or virtual), the percentage of reduction in cost for virtual courses starts on an interval between 2,5% and 6% depending on the incidence of costs of assets and facilities in the final budget of the action. This hypothesis assumes that the content of the course does not require learning *in situ*. For instance, a course *on the job* for digitization of herbarium sheets at the workstations located in an institution must be partly in person, though introductory themes and preliminary aspects can easily be presented online to analyse methodologies used, handling and management of the sheets. Following on with the premise of reduced cost of virtual course versus physical ones, such reduction will increase progressively and become larger as the course is repeated and even be maximized if the course is converted in a MOOC format (Massive Open Online course), that offers training content to an unlimited number of trainees, at anytime from anywhere.

[29] EUROSTAT website, inflation in the euro area, consulted in November 2022, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Inflation_in_the_euro_area

3.7 Decommissioning / Dismantling the RI

Decommissioning costs are meant to happen at the end of the life-cycle of the research infrastructure. The European Commission asks RIs to estimate them before commissioning the RI. They are particularly relevant for physical RIs such as nuclear facilities or those with materials that are expensive to dismantle.

In the case of DiSSCo, a virtual research infrastructure, it is not concerned with dismantling a physical infrastructure. Nevertheless, decommissioning to a wider extent also refers to human resources, employment and data. The impact on the RI employees (ending personal contracts) could be considered in that context as well as the way data is managed: it will be important to secure access to data beyond the RI life-cycle. DiSSCo will create and manage a great wealth of data stored in servers. Decommissioning costs could be seen as the cost of migrating the data when the system implemented by DiSSCo terminates because it is obsolete.

At this stage, it is reasonable to envisage that the duration of the DiSSCo decommissioning would last between one and two years. It could include the costs to redistribute the activities to the members. In addition, it could also include the costs to end contracts and agreements with partners. If we estimate the cost to run the RI and its services at around €M 2 per year, it is possible to give a rough estimation of the decommissioning costs. In the hypothesis that decommissioning will last 1 year, it would cost €M 2. In the hypothesis it will last 2 years, it will cost €M 4.

These figures are not set in stone and it is hard to find documentation on decommissioning virtual distributed research infrastructures. As we know that the actions implemented by the RI will evolve over time and that the end of the RI is highly unpredictable, it is the only figure WP4 can provide.

CONCLUSION

Deliverable 4.1 – DiSSCo Cost Book – aimed at calculating the costs to build, to operate and to dismantle the DiSSCo Research Infrastructure. To do so, WP4 needed to clearly understand and, eventually, foresee how the RI is going to work, the actions that are going to be implemented, the distribution of its resources, etc. These questions were raised during the period in which the RI was still under its preparatory phase, and therefore not all the responses were found as often the decisions were still pending.

In that sense, the first understanding is that what characterizes a distributed research infrastructure like DiSSCo is its decentralization, its potential of evolution according to funding opportunities and the willingness of its members to increase their participation in the RI. At a very early stage, there is a lot of room for future strategic development.

WP4 work on cost calculation then first relies on the four strategic pillars delimited for DiSSCo: digitisation, access, capacity building and e-services. They encompass a wide range of activities which are at this stage implemented by DiSSCo member institutions who preserve Natural Science Collections, share data and provide services. The goal was then to find a way to encourage them to quantify the costs associated. In this line, WP4 developed a methodology accessible to as many people as possible and shareable despite cultural and language differences. The objective of this methodology was to gather costs based on the same variables. Behind that strategy, the vision was to orientate the way partners estimate their costs and to reduce the lack of transparency from cost shared at European level.

The results of this strategy are partial: 27 institutions out of the 170 targeted shared data with WP4. Nevertheless, the main results are that, in proportion to the results received, annually €M 290 are spent to provide services associated with Natural Science Collections. With the same data collected, except from NHMW, is observed that the cost per hour, or the cost per services are in comparable price categories. The methodology developed by WP4 can then guide DiSSCo future teams to develop prices for the services of the RI. These results involve the related DiSSCo future activities implemented by its potential future nodes: the decentralized costs.

Concerning the centralized costs, the so-called “DiSSCo Central Hub Office” will be an interface between the member-institutions and users (industry, researchers, public institutions, etc.). The DiSSCo Central Hub office, according to the first calculation, would annually cost €M 1.4. This cost can change over time. Often, research infrastructure actions grow and initial budgets are lower at the beginning. In that context, the DiSSCo construction phase will be a strategic period to complete the development of the DiSSCo IT infrastructure and e-services. This period of time could cost around €M 2.2 (excluding the cost of the Central Hub office). The annual cost to maintain and run DiSSCo IT infrastructure could be around €M 1.2. It will participate in the production of Findable, Accessible, Interoperable, Reusable (FAIR) NSC data.

This framework provides an overview on how much the DiSSCo Research Infrastructure would cost. Here what is maybe most important, beyond the figures calculated, are the cost units defined. Whether for the cost areas or the nature of the costs, it is a framework that can be used for future cost calculation exercises in the field of Natural Science collections.

REFERENCES

Alex R Hardisty, Elizabeth R Ellwood, Gil Nelson, Breda Zimkus, Jutta Buschbom, Wouter Addink, Richard K Rabeler, John Bates, Andrew Bentley, José A B Fortes, Sara Hansen, James A Macklin, Austin R Mast, Joseph T Miller, Anna K Monfils, Deborah L Paul, Elycia Wallis, Michael Webster, Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet, *BioScience*, 2022;, biac060, https://doi.org/10.1093/biosci/biac060

Ambrosio J. How to build out your DevOps team, online, published on January 25 2022, consulted in June 2022. URL: https://about.gitlab.com/blog/2022/01/25/how-to-build-out-your-devops-team/

Bouquin, H., « Gestion financière - Comptabilité de gestion », Encyclopædia Universalis [en ligne], consulted on June 21st, 2022. URL : http://www.universalis-edu.com/encyclopedie/gestion-financiere-comptabilite-de-gestion/

Centre for Industrial Studies (CSIL) (2019) Guidelines on cost estimation of research infrastructures European Strategy Forum on Research Infrastructures (ESFRI). Accessed on 2021-09-21.

CIGREF, Modèle d'analyse et de benchmarking des coûts informatiques. En ligne, mise à jour de juillet 2018, consulted on June 21st, 2022. URL: https://www.cigref.fr/wp/wp-content/uploads/2018/07/2018-Modele_d_analyse_et_de_benchmarking_des_couts_informatiques-Cigref-mise-a-jour-2018_web.pdf

ERIC Forum Policy Brief, Funding Models for Access to ERIC Multinational/Transnational Services; accessed on October 2022, https://www.eric-forum.eu/wp-content/uploads/2020/09/ERIC-Forum_Policy-Brief.pdf

ESFRI methodology: https://www.esfri.eu/sites/default/files/StR-ESFRI2_STUDY_RIs_COST_ESTIMATION.pdf

EUR-Lex - 52021XC1213(01) - EN https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.C_.2021.501.01.0011.01.ENG&toc=OJ%3AC%3A2021%3A501%3AFULL - accessed on November 14th 2022

EUROSTAT website, inflation in the euro area, consulted in November 2022, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Inflation_in_the_euro_area

Hardisty A, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2021) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7: e67379. https://doi.org/10.3897/rio.7.e67379

Hardisty A, Livermore L, Walton S, Woodburn M, Hardy H (2020) Costbook of the digitisation infrastructure of DiSSCo. Research Ideas and Outcomes 6: e58915. https://doi.org/10.3897/rio.6.e58915

Hardisty A, Livermore L, Walton S, Woodburn M, Hardy H (2020). Costbook of the digitization infrastructure of DiSSCo.

Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalga A, Paul DL, 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: e54280. https://doi.org/10.3897/rio.6.e54280

Horizon Europe – Work Programme 2021-2022, Marie Sklodowska- Curie Actions, European Commission DecisionC (2022) 2975 of 10 May 2022, URL: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-2-msca-actions_horizon-2021-2022_en.pdf, accessed on 23/06/3022

Islam S, Weiland C, Addink W (2022) From data pipelines to FAIR data infrastructures: A vision for the new horizons of bio- and geodiversity data for scientific research. Research Ideas and Outcomes 8: e93816. https://doi.org/10.3897/rio.8.e93816

Islam, S., Hardisty, A., Addink, W., Weiland, C. and Glöckler, F., 2020. Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections. Data Science Journal, 19(1) DOI: http://doi.org/10.5334/dsj-2020-050

Landel, S., Guiraud, M. (2022). IT infrastructure and data preservation indicators for DiSSCo RI Cost Book. DiSSCo Prepare WP4 – Ms4.7.

Leeflang, Sam, Weiland, Claus, Grieb, Jonas, Dillen, Mathias, Islam, Sharif, Fichtmueller, David, Addink, Wouter, & Haston, Elspeth. (2022). DiSSCo Prepare D6.2 Implementation and construction plan of the DiSSCo core architecture (1.2). Zenodo. https://doi.org/10.5281/zenodo.6832200

Pêcheux F. « Cloud computing ou informatique dans les nuages », Encyclopædia Universalis [en ligne], consulted on June 22nd, 2022. URL : http://www.universalis-edu.com/encyclopedie/cloud-computing-informatique-dans-les-nuages/

Petersen M. et al. (2022): DiSSCo Prepare Milestone report MS5.2 “Implementation of concepts for sustainability of services, CMS, and overall TRL”. https://doi.org/10.34960/1xr6-hr45
Sotiriou, S (et al.), 2022. Deliverable D2.1 "DiSSCo Training Strategy"

Synthesys+ - D2.3 – catalogue and recommendations for the development of a proactive, efficient and evolving DiSSCo training programme.

Walton S, Livermore L, Dillen M, De Smedt S, Groom Q, Koivunen A, Phillips S (2020) A cost analysis of transcription systems. Research Ideas and Outcomes 6: e56211. https://doi.org/10.3897/rio.6.e56211

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
Wittenburg P, Strawn G. Revolutions Take Time. Information. 2021; 12(11):472. https://doi.org/10.3390/info12110472

Appendix 1: DiSSCo financial glossary

Accounting entity:	An entity based on the principle of subsidiarity (subsidiarity is a principle of social organization that holds that social and political issues should be dealt with at the most immediate (or local) level that is consistent with their resolution). The organisational unit in your institution that is most able to measure the costs of a specific DiSSCo area or scientific domain.
Amortization of DiSSCo data:	Amortization is the process of allocating the costs of an intangible asset such as data over time. The purpose is to match the costs of creating and maintaining data to the value earned from using that data. Or to put it another way, to ensure that expenses are not incurred in maintaining data with no useful value
Analytical accounting system:	Is a tool that helps you to analyse, interpret, and create reports based on your company's chart of accounts. With Analytical Accounting, you can: set up unlimited analysis dimensions; enter analysis information for a group of analysis dimensions; create budgets using the analysis dimensions you've set up; perform comprehensive reporting by exporting analysis queries to Microsoft Excel.
Capital cost / expenditure / investments:	Capital costs are fixed, one-time costs incurred on the purchase of equipment, buildings, construction to be used for digitisation. If in doubt about what to count as capital, a general rule is that if an asset has a useful life of more than one year, it is a capital cost.
Construction phase:	Implementation of national / regional investment plans for infrastructure upgrades and large-scale digitisation programmes; application of joint DiSSCo programmes and policies, quality control and risk management; establishment of regional/ thematic hubs; active membership; construction of the DiSSCo Hub (including all services).
Decommissioning costs:	When there are dangerous materials or waste to be disposed of after the shutdown of the facility. To officially take a factory or other industrial building out of use and make the area safe. For DiSSCo, these costs will mainly concern human resources, employment and data.
Depreciation of equipment:	Depreciation is the process of allocating the capital costs of a tangible asset (such as digitisation equipment or storage systems) over time. It's a measure of how much of the value of an asset has been consumed to a point in time (usually, the end of an accounting period). Depreciation is well understood and, especially for IT infrastructure, is typically allocated over 3 or 4 years using a straight-line method (i.e. the same amount in each year).
Direct costs:	Addition of staff costs, operating cost and capital cost directly connected to DiSSCo areas. For instance, time spent by a curator to preserve specimens is a direct cost. On the contrary, time spent by HR to manage the wage of this curator is an indirect cost.

Full economic cost (FEC):	Total cost of the employee for the employer. Considers all charges associated with employees (includes sums paid under the employee savings scheme, bonuses, employer costs). It does not consider overheads related to employees.
Full time equivalents (FTE):	corresponds to an activity carried out on the basis of a full time position up to legal duration. This legal duration may vary from country to country. In France, one FTE corresponds to 1,607 hours of annual work.
Indirect costs / support costs / overheads:	Indirect costs are those costs covered by institutions that are involved in the effective functioning of projects and are common to the whole institution. Two types of indirect costs have been identified: (i) Indirect administrative costs: human resources, legal services, finance department, presidency, IT services, etc. These costs are common and contribute to the efficiency of the services provided by the staff; (ii)Indirect building costs: technical maintenance of buildings and fluid consumption. These costs are essential to provide working space for the staff working on the project.
Major upgrade:	extraordinary maintenance and major upgrades are investment costs which occur during the operational phase and are related to the modernization and expansion of the facility. They are interventions which modify the performance in a structural way and produce effects beyond the financial year(s) in which they take place.
Operating costs:	Expenditure on a non-durable good. Operating costs can be fixed or variable. Typical operating costs include: rent of buildings or sheds, rental of machinery; personnel; ordinary maintenance and repair of assets; utilities (consumption of raw materials, fuel, energy) and consumables; users support, services purchased from third parties. **Within the Cost Book, we make the distinction between staff costs and expenditure on non-durable goods. Operating costs only correspond to expenditure on a non-durable good.** Example: an annual contract for maintenance worth €10,000 is an operating cost as it is not a durable good. Example 2: Purchase of a pack of notebooks worth €50 = operating cost.
Staff costs:	Expenditure incurred for staff time used to deliver projects. Example –Annual salary, national insurance, pension contributions, employer's contributions for NI and pension, any other contractual payments included in the employee contract

Appendix 2 – DiSSCo perimeter glossary

Asset maintenance:	conservation and data curation. It covers the renovation of collections (relates to work sites, relocation of collections, restoration of a room, etc.), the preventive and curative preservation (day-to-day maintenance of collections to ensure their long-term preservation) and data curation activities (time spent to maintain, manually or computationally, information about natural science collections. It involves updates of digitised information).
Collection analyses services:	produce data that would not be visible to the naked eye. They require a physical intervention on a specimen (sampling, imaging, etc.) and the use of specific analytical facilities (laboratory, microscope, etc.).
Consulting services / expertise:	corresponds to any requests that rely on giving expert advice, for example determining taxonomy. There is a lot of collection information which is not available through an externally accessible database: collections are not fully digitized, both quantitatively and qualitatively. Some of this information often relies upon the knowledge and experience of collection managers and curators. Consulting services can be the direct interaction with collection managers and curators.
Cost Book areas:	Correspond to activities implemented within natural history institutions. Are grouped in association with services for DiSSCo. Are meant to be homogeneous and coherent.
Data curation:	correspond to the time spent to enter, manually or computationally, information about natural history collections. This involves both initial registration of information and updates. Data curation is directly related to the Mids level of the specimen (see D2.3 - Design of a Collection digitisation Dashboard - ICEDIG). This cost area includes staffing costs associated with improving data curation activities should be classified here or under IT tools (e.g. time on external projects like TDWG collections descriptions interest group, technical staff time on internal data curation projects such as Join the Dots).
Digitisation:	Two possibilities: mass digitisation and routine digitisation. Digitisation includes the process of converting analog information about physical specimens to digital format which includes electronic text, images and other representations.
Digitised specimen:	Taking pictures of specimens and databasing. This can range from a very elaborate digitisation method to simple photography.
DiSSCo Perimeter	The DiSSCo perimeter can be represented with concentric circles. At the centre there is a hub which coordinates demand and service provision. Around this hub, there is the ERIC perimeter (first circle): it can encompass other facilities than the hub - in the scenario, Service Level Agreements (SLAs) are signed with institutions who become service providers on behalf of the ERIC. Finally, there is the second circle which is the Research Infrastructure. Inside it are the institutions who preserve NSCs and provide services and data.

Loans of collections:	time accounted for in DiSSCo excludes time spent on loans for exhibition but includes time starting with decision-making committees for these requests, to finding the requested specimens, assessing their condition, sampling fragments if requested, packaging and customs formalities.
Mass digitisation:	is the implementation of a workflow to digitise a large number of specimens in a limited time. It includes the preparatory phase of digitisation up to the production and publication of the data. Indicative threshold: more than 5 000 digitised specimens in a year (up to tens or hundreds of thousands in some cases). Digitisation does not automatically mean imaging, it can only be databasing. By mass digitization, we mean digitizing entire collections or their major distinct parts at industrial scale (i.e, millions of objects annually at low cost), characterised by improved workflows, technological and procedural framework based on automation (both hardware and software) and enrichment (link-building).
National node network activity:	Corresponds to coordination activities of the national node. They range from administrative management of the node, to promotion of the DiSSCo RI, as well as team engagement activities. These engagement activities are intended for local nodes who are represented by the national node who have signed the DiSSCo MoU. It has no relation with the general governance of your institution.
Physical access to collections:	the time accounted for starts with decision-making committees for these requests, to prepare the specimens for consultation, take care of the material, organization of the visit and deal with the administrative requirement for hosting the visitor.
Preservation / conservation costs:	DiSSCo provides data and images from physical collections held by its members. This raises the issue of how to account for preservation costs. Collections are irreplaceable archives of nature that can be reused almost indefinitely. The cost of preserving these archives is borne by the institutions, but it is conceivable that users could be asked to contribute to their preservation. Conservation activities start from the moment specimens enter collections. This excludes specimen acquisition.
Preventive and curative preservation	corresponds to day-to-day maintenance of collections to ensure their long-term preservation.
Renovation of the collections:	relates to worksites, relocation of collections, restoration of a room. Has an unusual nature.

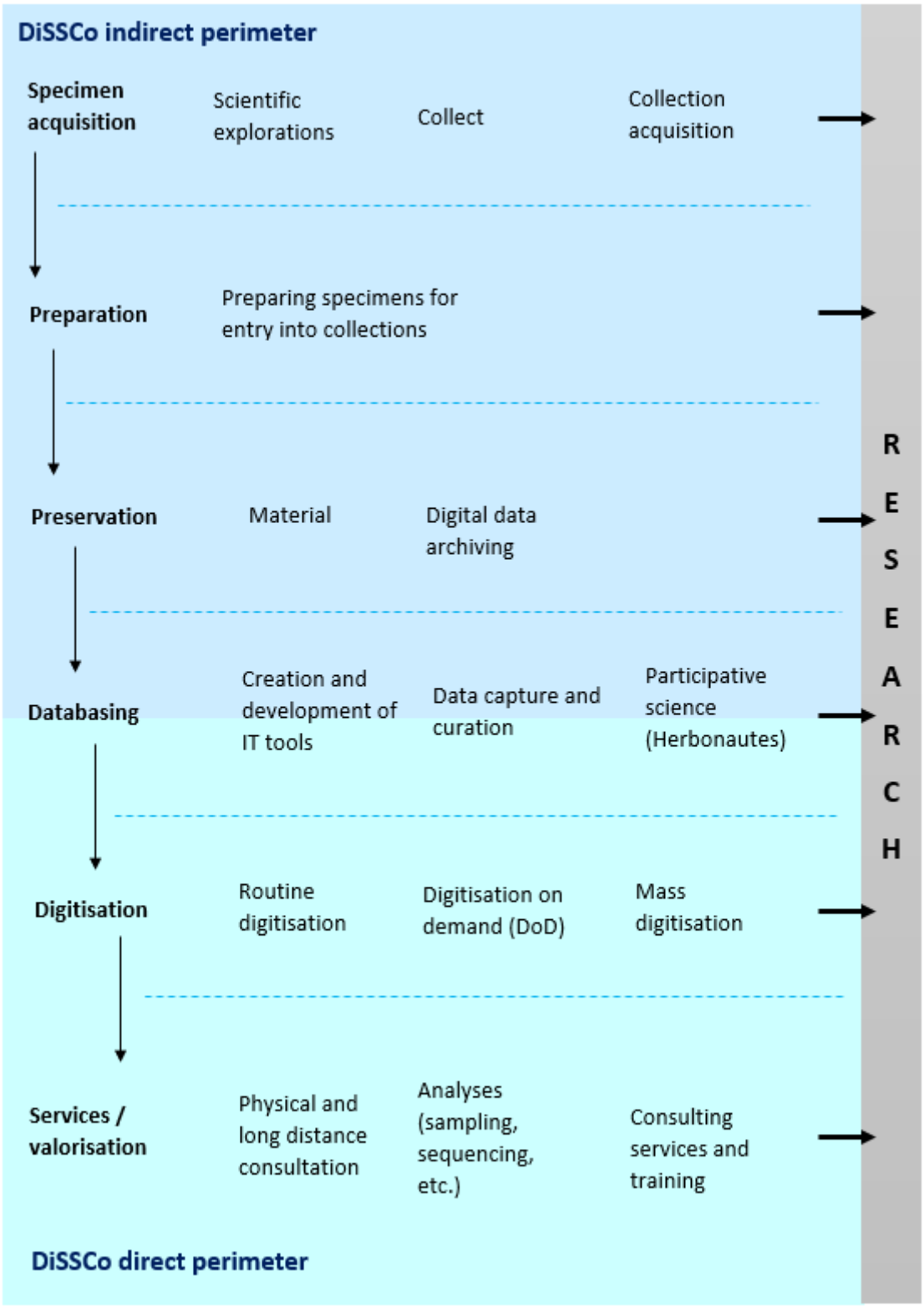
Running IT infrastructure:	The main initial service of DiSSCo to external users is likely to be providing access to data, images and other media that have been already produced by members of DiSSCo. This includes: **Data management infrastructure:** the use of equipment to manage the data that will be shared with DiSSCo. This requires staff time, computers and appropriate IT systems. This sub-area includes the design of databases and other software to use the data. **Data storage:** corresponds to servers and other equipment, software to manage data and other media attached to servers. This category corresponds to short-term backup storage. **Digital preservation:** corresponds to formal means of ensuring permanent access to digital information. Archiving involves deep storage with less frequent retrieval. Preservation is bound up with particular technology - for example, automatically checking that image quality has not degraded (bit degradation, etc.) and updating data formats. It involves software to monitor files and image formats.
Small set / Routine digitisation:	In contrast to mass digitisation, this digitisation is done on a regular basis and/or on demand, and involves only a smaller number of specimens at a time. Indicative threshold: less than 5 000 digitised specimens in a year. Digitisation does not automatically mean imaging, it can only be databasing.
Specimen acquisition:	Scientific explorations to collect new specimens that will become natural history collections.
Training	courses where the participation of experts from Natural Science institutions is crucial. Also important is the tooling-up of the scientific community in their different areas of expertise. The courses can address the needs of staff in different stages of their careers, ranging from early-career investigators, young professionals to advanced more experienced senior staff members. They both target RI members and non-members. [30]

[30] Definition from Synthesys+ - D2.3 – catalogue and recommendations for the development of a proactive, efficient and evolving DiSSCo training programme.

Appendix 3: DiSSCo general glossary

Central hub:	The infrastructure of integrating services, information technology components (hardware and software), human resources, organisational activities, governance, financial and legal arrangements that collectively have the effect of unifying natural science collection through a holistic approach towards digitization of and access to the data bound up in those collections.
DiSSCo facility:	The geographically distributed collection-holding organisation(s) (i.e., natural science / history collection(s)) and related third-party organisations that deliver data and expertise to the DiSSCo Hub infrastructure, and which can be accessed by users via the DiSSCo Hub infrastructure. A special kind of DiSSCo facility could be a DiSSCo Centre of Excellence (DCE), specialised in one or more of researching, innovating, developing and operating / performing techniques and/or processes of digitization or other related facets, and disseminating information on same/
DiSSCo National Consortium:	a consortium of two or more Natural Science Collection-Based Organisations from the same country, formulated to pursue, among other things, the representation of its constituent organisations at the DiSSco bodies and contributes to the set-up and operation of the DiSSCo natioanl node. A National Consortium may or may not have a separate legal personality.
Distributed RIs:	is «a network of distributed resources» and consists of a Central Hub (i.e. coordination secretariat) and interlinked National Nodes, which can further coordinate local nodes. A first important distinction is between distributed RIs (which, according to the ESFRI roadmap, need – among other requirements – to be identified by a unique name, legal statutes, and governance structure) and a coordinated research network, which is instead the collaboration of fully independent research performing organizations.
Local entities:	Other institutions part of DiSSCo but who are not national nodes.
National Node representative:	Institution representing the (informal or formal) group gathering signatories to DiSSCo MoU in one country with representation agreed among the entities to participate in DiSSCo Prepare project and other DiSSCo-related consultation fora.
Scientific domains:	Same as for CETAF passport Collections and Collections Registry: • Algae-Fungi-Plants; • Anthropology; • Data; • Extra-terrestrial; • Geology; • Heritage (Library, archives); • Zoology invertebrates; • Microorganisms; • DNA / RNA / Tissue / Biobank; • Palaeontology; • Zoology vertebrates.

Appendix 4: DiSSCo RI perimeter



Appendix 5: Detailed DiSSCo Central Hub Office Costs (Construction phase)

DiSSCo Central hub office - Construction phase				
	Cost per unit	Type of units (day / month / package)	Number of units	Total
Coordination				**506 950 €**
Director general	222 555 €	/ year	1	222 555 €
Service director	*173 851 €*	/ year	1	173 851 €
Change / innovation manager	*110 544 €*	/ year	1	110 544 €
Administration				**249 710 €**
Financial administrator	*166 840 €*	/ year	1	166 840 €
Administrator / secretary	*82 870 €*	/ year	1	82 870 €
Premises / office management				**143 000 €**
Furniture / office supplies	5 000 €	/ year	1	5 000 €
Insurance	5 000 €	/ year	1	5 000 €
Rent and charges	100 000 €	/ year	1	100 000 €
Internet / mobile phones	5 000 €	/ year	1	5 000 €
IT equipment for the team	2 000 €	package/pers	9	18 000 €
Other equipment / consumables	10 000 €	/ year	1	10 000 €
Communication				**175 544 €**
Equipment for communication	10 000 €	/ year	1	10 000 €
Brochures	15 000 €	/ year	1	15 000 €
Events	20 000 €	/ year	1	20 000 €
Communication officer	110 544 €	/ year	1	110 544 €
Travel	20 000 €	/ year	1	20 000 €
Outsourcing and services				**43 000 €**
Advertising	2 000 €	/ year	1	2 000 €
Bank account charges, credit card charges	1 000 €	/ year	1	1 000 €
External auditor	4 000 €	/ year	1	4 000 €
External legal consulting / advising	15 000 €	/ year	1	15 000 €
Translation	1 000 €	/ year	1	1 000 €
Website	20 000 €	/ year	1	20 000 €
Other		package		0 €
Other expenses				**7 000 €**
Training of DiSSCo personnel	7 000 €	/ year	1	7 000 €
Total direct costs				**1 125 204 €**
Overheads (25%)				**270 551 €**
			Total costs	**1 395 755 €**

Appendix 6: Detailed DiSSCo Central Hub Office Costs (Operation phase)

DiSSCo Central Hub Office - Operation phase				
	Cost per unit	Type of units (day / month / package)	Number of units	Total
Coordination				**525 092 €**
Director general	228 746 €	/ year	1	228 746 €
Service director	*181 156 €*	/ year	1	181 156 €
Change / innovation manager	*115 189 €*	/ year	1	115 189 €
Administration				**259 027 €**
Financial administrator	*173 851 €*	/ year	1	173 851 €
Administrator / secretary	*85 176 €*	/ year	1	85 176 €
Premises / office management				**129 500 €**
Furnitures / office supplies	5 000 €	/ year	1	5 000 €
Insurance	5 000 €	/ year	1	5 000 €
Rent and charges	100 000 €	/ year	1	100 000 €
Internet / mobile phones	5 000 €	/ year	1	5 000 €
IT equipment for the team	500 €	package/pers	9	4 500 €
Other equipment / consumables	10 000 €	/ year	1	10 000 €
Communication				**180 189 €**
Equipment for communication	10 000 €	/ year	1	10 000 €
Brochures	15 000 €	/ year	1	15 000 €
Events	20 000 €	/ year	1	20 000 €
Communication officer	115 189 €	/ year	1	115 189 €
Travel	20 000 €	/ year	1	20 000 €
Outsourcing and services				**43 000 €**
Advertising	2 000 €	/ year	1	2 000 €
Bank account charges, credit card charges	1 000 €	/ year	1	1 000 €
External auditor	4 000 €	/ year	1	4 000 €
External legal consulting / advising	15 000 €	/ year	1	15 000 €
Translation	1 000 €	/ year	1	1 000 €
Website	20 000 €	/ year	1	20 000 €
Other		package		0 €
Other expenses				**7 000 €**
Training of DiSSCo personnel	7 000 €	/ year	1	7 000 €
Total direct costs				**1 143 808 €**
Overheads (25%)				**275 202 €**
			Total costs	**1 419 010 €**

Appendix 7: Detailed DiSSCo Central Hub IT costs (construction phase)

DiSSCo Central Hub IT - Construction costs to finalise the IT system on several years					
Type of Costs	Type of Units	Average Cost Per Unit (€)	Number of Units Required to fully develop service	Total Cost (€)	Uncertainty level
Staff cost				***1 340 533***	
Lead Architect	*FTE*	*90 428*	*2*	*173 321*	*4*
Lead Developer	*FTE*	*90 428*	*4*	*361 712*	*4*
Lead Operations/Systems Engineer	*FTE*	*79 923*	*2*	*159 847*	*4*
Developer	*FTE*	*79 923*	*6*	*439 579*	*4*
Product Owner/Project Manager	*FTE*	*90 428*	*1*	*90 428*	*4*
Data administrator	*FTE*	*70 088*	*2*	*115 646*	*4*
Outsourcing cost				**432 000**	
Subcontractor	*Cost per Day*	*1 000*	*180*	*180 000*	*3*
User interface design	*Cost per Day*	*1 000*	*37*	*37 000*	*4*
Membership fees (ORCID, Datacite, TDWG)	*Cost per year*	*15 000*	*3*	*45 000*	*2*
DOIs	*Cost per year*	*30 000*	*2*	*60 000*	*4*
External services (Dina, geocase, catalogue of life, etc.)	*Cost per year*	*50 000*	*2*	*100 000*	*4*
Software Licences	*Cost per year*	*5 000*	*2*	*10 000*	*2*
Cloud computing / hosting costs				**102 800**	**4**
Test environment	AWS/month	900	24	21 600	2
Acceptance environment (sandbox.dissco.tech)	AWS/month	1 600	36	57 600	2
Production environment	AWS/month	1 950	*12*	23 400	4
Wordpress sites	hosting per year	100	*2*	200	1
Total direct costs				**1 875 333**	
Indirect costs				**335 133**	
Total costs per year				**2 210 466**	

Appendix 8: Detailed Central Hub IT costs (operation phase)

DiSSCo Central Hub IT - Operation costs per year					
Type of Costs	**Type of Units**	**Average Cost Per Unit (€)**	**Number of Units Required to fully develop service**	**Total Cost (€)**	**Uncertainty level**
Staff cost			***7,8***	***633 995***	
Lead Architect	*FTE*	*90 428*	*0*	18 086	*4*
Lead Developer	*FTE*	*90 428*	*1*	90 428	*4*
Lead Operations/Systems Engineer	*FTE*	*79 923*	*1*	39 962	*4*
Developer (backend and frontend)	*FTE*	*79 923*	*3*	239 770	*4*
Product Owner/Project Manager	*FTE*	*90 428*	*1*	126 599	*4*
Data administrator	*FTE*	*70 088*	*2*	119 150	*4*
Outsourcing cost			**25**	**270 000**	
Subcontractor	*Cost per day*	*1 000*	*20*	20 000	*4*
				0	
Membership fees	*Cost per year*	*15 000*	*1*	15 000	*2*
DOIs	*Cost per year*	*30 000*	*1*	30 000	*4*
External services (Dina, geocase, catalogue of life, etc.)	*Cost per year*	*50 000*	*1*	50 000	
Software Licences	Annual licence	5 000	1	5 000	*2*
Membership fee Catalogue of Life (COL)	*Cost per year*	*150 000*	*1*	*150 000*	*3*
Cloud computing / hosting costs			**38**	**53 600**	
Test environment	AWS/month	900	12	10 800	2
Acceptance environment (sandbox.dissco.tech)	AWS/month	1 600	12	19 200	2
Production environment	AWS/month	1 950	12	23 400	4
Wordpress sites	hosting per year	100	*2*	200	
Total direct costs				**957 595**	
Upgrades				**95 760**	
Indirect costs				**158 499**	
Total annual costs				**1 211 854**	

Appendix 9 – Detailed summary DiSSCo Central Hub IT costs

		E-Service	E-Service	E-Service	E-Service	E-Service	Total e-services	Core services	Core services	Core services	Core services	Core services	Total Core services
		ELViS	CDD	SDR	Knowledgebase	Helpdesk		DO Repo	AAI	UCAS	PID	DMF	
Construction phase = total costs to finalise the IT System													
TOTAL direct costs (K€)	1 875	**310**	**55**	**125**	**57**	**0**	**548**	**504**	**135**	**279**	**106**	**90**	1 115
Staff (K€)	1 341	130	40	125	57	0	353	493	75	268	61	90	988
Outsourcing (K€)	432	180	15	0	0	0	195	11	60	11	45	0	127
Cloud computing (K€)	103	This total cost is distributed among the different listed components											
Indirect costs (K€) (25% of staff costs)	335	This total cost is distributed among the different listed components											
Total construction costs (K€)	2 210												
Operation phase													
TOTAL direct costs (K€)	958	**100**	**7**	**9**	**7**	**12**	**135**	**199**	**22**	**112**	**190**	**0**	524
Staff (K€)	634	80	7	9	7	7	110	199	22	112	190	0	524
Outsourcing (K€)	270	20	0	0	0	5	25	0	0	0	0	0	0
Cloud computing (K€)	54	This total cost is distributed among the different listed components											
Upgrades (K€)	96	This total cost is distributed among the different listed components											
Indirect costs (K€) (25% of staff costs)	158	This total cost is distributed among the different listed components											
Total operation costs (K€)	1 212												

Appendix 10: Country correction coefficients applicable to the remuneration and pensions of officials and other servants of the EU[31]

	Remuneration	Transfer	Pension
Country / Place[32]	01/07/2021	01/01/2022	01/07/2021
Austria	109,6	113,5	113,5
Bulgaria	61,7	58,6	
Croatia	78,3	69	
Cyprus	82,2	84,6	
Czechia	88,1	75,1	
Denmark	134,2	136,9	136,9
Estonia	86,3	90,2	
Finland	118,6	121,4	121,4
France	119,9	111,2	111,2
Germany	101,4	101,3	101,3
Greece	85,2	82,3	
Hungary	76,1	63,6	
Ireland	133,6	125,4	125,4
Italy	95,2	97,1	
Karlsruhe	96,9		
Latvia	80	74,6	
Lithuania	80,1	70,2	
Malta	94	99	
Munich	113,4		
Netherlands	111,4	111,3	111,3
Poland	70,6	61,3	
Portugal	91,4	87	
Romania	68,5	57,9	
Slovakia	79,9	77,7	
Slovenia	84,9	81,9	
Spain	96,3	93,3	
Sweden	130,3	120	120
United Kingdom	n/a	n/a	128,5

[31] EUR-Lex - 52021XC1213(01) - EN https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.C_.2021.501.01.0011.01.ENG&toc=OJ%3AC%3A2021%3A501%3AFULL - accessed on November 14th 2022

[32] Belgium is equal to 100 and does not have a CCC

Appendix 11: List of institutions who participated in the DiSSCo Cost Book study

#	Country code	List of institutions	Acronyms	Test 1 (2020)	Test 2 (2021)	Final (2022)
1	AT	Naturhistorisches Museum Vienne	NHMW			X
2	BE	Meise Botanic Garden	Meise		X	X
3	BE	Royal Belgian Institute of Natural Sciences (Brussels)	RBINS		X	
4	DE	Botanischer Garten und Botanisches Museum (Berlin)	BGBM			X
5	DE	Museum für Naturkunde Berlin	MfN			X
6	DE	Naturkundemuseum Stuttgart	SMNS			X
7	DE	Senckenberg Gesellschaft für Naturforschung (Frankfurt)	SGN		X	
8	DK	University of Copenhagen, Natural History Museum of Denmark	NHMD			X
9	EE	Univeristy of Tartu National History Museum and Botanical Garden	Tartu			X
10	FR	Centre de coopération internationale en recherche agronomique pour le développement	CIRAD			X
11	FR	Institut de recherche pour le développement	IRD			X
12	FR	Muséum d'histoire naturelle de Marseille	MHNM			X
13	FR	Muséum d'histoire naturelle de Nice	MNHNice			X
14	FR	Muséum d'histoire naturelle de La Rochelle	MNHNLR			X
15	FR	Muséum national d'Histoire naturelle (Paris)	MNHN	X		
16	FR	Université Claude Bernard Lyon 1	Lyon 1			X
17	FR	Université de Bourgogne	UB			X
18	FR	Université de Rennes 1	Rennes			X
19	FR	Université de Montpellier	Montpellier			X
20	IT	Università degli Studi di Firenze	UNIFI			X
21	NL	Naturalis Biodiversity Centre (Leiden)	Naturalis			X
22	NL	Natuurhistorisch Museum Maastricht	NHM NL			X
23	NL	Natuurmuseum Brabant	NMB			X
24	NL	Royal Netherlands Institute for Sea Research	NIOZ			X
25	NL	Teylers Museum	Teylers			X
26	NL	Natuurmuseum Fryslân	NMF			X
27	NL	Natural History Museum Rotterdam	NMR			X
28	UK	Natural History Museum (London)	NHM		X	X

Appendix 12: DiSSCo Cost Book recommendations

N#	DiSSCo Cost Book Recommendations
1	From an economic viewpoint, a good understanding of the costs related to the activities implemented at institution-level can improve their efficiency. As good practice, ongoing recording of the costs can help to develop more cost-effective projects.
2	When implementing a costing methodology, it is important to collect the annual expenditure and to ask for the number of services provided per year (number of loans, number of visits, number of specimens digitised, etc.). With that information, it is possible to calculate the unit cost of the services. It is an incentive for the partners as it is also a tool to establish a pricing system.
3	When sharing a cost methodology, it is possible to consider that the smaller the impact a decision might have on the final cost calculation, the more it is possible to exclude it, instead of wasting people's time.
4	In the case of the DiSSCo perimeter, the difficulty is that it is transversal across different activities of the institutions. It cannot be copied on a pre-existing perimeter. In that context, traditional accounting systems, with pre-existing cost categories, cannot provide all the information needed. It requires human beings who will look for the information and analyse it in accordance with the predefined scope.
5	Information on costs can be in the hands of a few contact persons in position to have a good understanding of the staff activities, annual operating and investment costs. Generally, it is the managers and directors of departments. It can also be administrative managers, with information on the department costs. In order to avoid contacting too many people, it is important to identify individuals who can be pivotal in the Cost Book process.
6	In order to make sure that the data collected is coherent, there is a need to define coherent categories and areas of costs.
7	The methodology should be short, without too much text, and if possible with images and diagrams. It allows for a rapid understanding of the concepts used and, in case there is a question, people can contact the team.
8	In order to get the same data from all partners, the methodology shall not be changed over the course of implementation.
9	Regularly send reminders. If the methodology is sent only once, it is possible that nobody will follow it. In order to be credible, it is important to repeat the request and therefore to have rather long deadlines.
10	Leave a room for uncertainty: it allows people participating in the survey to self-assess the level of uncertainty of their responses. These levels are directly connected to formulas which can calculate a range between which the accurate costs are.
11	Direct costs correspond to staff costs, operating costs and investments directly connected to the project.

N#	DiSSCo Cost Book Recommendations
12	Staff cost is the most important type of costs in the DiSSCo context. It is the type of costs which requires the most attention.
13	Operating costs have a minor impact on the total costs. They are not always clearly identified. In the case of DiSSCo, it is possible to use large figures and a rough distribution within the different DiSSCo cost areas.
14	Capital costs / investment: use an average of the investments of the 5 years preceding the cost assessment. This would capture the variability of investment, including no investment in some years and high investment in others.
15	For indirect costs, use the flat rate proposed by the EU for these costs: 25% on direct costs. The 25% rate is linked to staff costs as sometimes operating costs and investments can be outsourced.
16	Important to value the maintenance / preservation of natural history collections. Option proposed: the maximum cost corresponds to the aggregation of all the costs provided and the asset maintenance proportional to its weight in relation to the AE total costs. It can be understood that without this maintenance, DiSSCo services would not exist as the collections would no longer be accessible.
17	If large equipment, machines, etc. are bought for the research infrastructure, the full price is part of the construction phase costs, and renewal could be considered as a replacement cost . Depreciation costs are not considered according to ESFRI methodology. Still, in the case where a large piece of equipment is bought, it is important to have information on its maintenance costs and an estimation of its duration: 5,10,20, … years
18	The purchase of a physical asset can lead to decommissioning costs: the cost to dismantle the asset has to be considered.
19	When it comes to the DiSSCo perimeter and more specifically its IT infrastructure, teams should consider both back office and front office.
20	To calculate the costs of DiSSCo RI, it is important to distinguish costs that will be invested to build the RI (construction phase) from the costs incurred to operate / run the RI (operation phase).
21	Make the distinction between the different components. It will allow the ERIC to plan its budget and increase its investments over time. The goal is then to estimate the distribution of the different cost units according to the different components of IT architecture. These results would help to distribute the investments required over time and, in the case where the services would be geographically distributed, among institutions.
22	Consider number of users when it comes to IT software applications: there are user thresholds above which the whole system should be redesigned to cope with capacity.
23	Consider the level of service the RI is engaged to provide, it could be more or less costly.
24	Consider the costs to train the service providers and users on the tools they will share and use: These costs can sound marginal when it comes to IT development but such tools are often

N#	DiSSCo Cost Book Recommendations
	essential. A RI like DiSSCo is nothing without its users and members. It requires a budget to engage and train them on the tools the RI proposes.
24	Consider major upgrades: updates are required at some regular time intervals and could represent a proportion of the costs connected with building the infrastructure (sometimes it is as expensive to upgrade a service as it is to build it).
24	To estimate training costs, it is possible to use the standard cost defined for the ERASMUS+ programme: € 80 per participant per day.
24	In the case of DiSSCo, a virtual research infrastructure, it is not concerned with dismantling a physical infrastructure. Nevertheless, decommissioning to a wider extent also refers to human resources, employment and data. The impact on the RI employees (ending personal contracts) could be considered in that context as well as the way data is managed: it will be important to secure access to data beyond the RI life-cycle. DiSSCo will create and manage a great wealth of data stored in servers. Decommissioning costs could be seen as the cost of migrating the data when the system implemented by DiSSCo terminates because it is obsolete.