

# DiSSCo RI: the Cost Book for DiSSCo

Salomé Landel<sup>‡</sup>, Michel Guiraud<sup>‡</sup>, Ana Casino<sup>§</sup>, Katharine Worley<sup>‡</sup>

<sup>‡</sup> Muséum national d'Histoire Naturelle, Paris, France  
<sup>§</sup> CETAF, Brussels, Belgium

Corresponding author: Salomé Landel ([salome.landel@mnhn.fr](mailto:salome.landel@mnhn.fr)), Michel Guiraud ([mguiraud@mnhn.fr](mailto:mguiraud@mnhn.fr)),  
Ana Casino ([ana.casino@cetaf.org](mailto:ana.casino@cetaf.org))

Reviewable v 1

Received: 21 Jan 2025 | Published: 21 Feb 2025

Citation: Landel S, Guiraud M, Casino A, Worley K (2025) DiSSCo RI: the Cost Book for DiSSCo. Research Ideas and Outcomes 11: e147364. <https://doi.org/10.3897/rio.11.e147364>

## 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 both a central team and a coordinated network responsible for supplying the infrastructure's services. The RI linked costs are spread all over Europe and connect 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 amongst a great number of partners, cost information is spread across centralised and decentralised areas.

The following article outlines the first "centralised" cost calculation exercise for DiSSCo and concludes that the DiSSCo Central Hub office would need a minimum annual budget of 1.4 million euros to be operational. This would not change between the construction and the operation phases. Furthermore, the DiSSCo Central Hub IT team would need a budget of 2.2 million euros to finalise all IT systems under the construction phase (if the construction phase lasts two years, it would cost around 1.1 million euros 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 1.2 million euros 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 is often the case, research infrastructures grow over time and the more DiSSCo becomes 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, we developed a cost calculation methodology that has been distributed amongst all the 170 DiSSCo partner institutions. Twenty-seven institutions responded to the exercise. This 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 basis and it opens the door for the pricing of DiSSCo services. Finally, such a methodology also aims at guaranteeing a fair service pricing, based on the same principles and variables.

## Keywords

cost calculation, construction phase, implementation phase, staff costs, operating costs, capital costs, costing methodology, Central Hub, cost per hour, Cost Book, Service provision, pricing, research infrastructure cost

## 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 advances 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 infrastructure aims to serve users across many disciplines. The NSCs contain data that, if extracted, could help in questions related to societal challenges such as global change, food security and health. Therefore, the business perimeter of DiSSCo is extensive, gathering many stakeholders in a process of transition from a fragmented landscape to an integrated research infrastructure.

As a distributed infrastructure, DiSSCo is composed of two elements: member institutions which provide the services; and a superstructure which monitors the RI and organises the relationships between the users and the members.

The DiSSCo infrastructure is also characterised by its members providing the same services; the distributed feature resides not in different services distributed across the RI, but in different collections and expertise that can be accessed using identical services. Therefore, approaching the costs of the infrastructure had to be the same across the different members in order to obtain a consistent evaluation.

The 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.)” (Cigref 2018, Centre for Industrial Studies and European Strategy Forum on Research Infrastructures 2019, Cigref 2018). For DiSSCo, this involves studying the centralised and decentralised costs of the RI. The cost book 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 detailed analysis of all the costs related to NSCs and identifies methods of calculating costs in a way that is achievable by the project members.

The aim of the cost book is a full cost analysis of DiSSCo. Such a task first requires establishing the perimeter of the RI and its associated resources. As DiSSCo services are spread amongst its 170-member institutions, cost calculation needs a methodology allowing for the participation of a wide range of stakeholders, with different levels of accounting skills. Indeed, the DiSSCo perimeter crosses some established boundaries: 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. This recognises the differentiating variable of price differences amongst countries, based on purchasing power parity.

## Project context

This project report was originally written as a formal deliverable (D4.1) of the DiSSCo Prepare project (Landel et al. 2023). As such, the document was reviewed by project partners and submitted to the European Commission. Some minor changes have been made to the original deliverable in order to render this version suitable for publishing; however, the authors consider this as the definitive version of the report.

The following text is the formal task description (Task 4.1) from the DiSSCo Prepare project's Description of work:

*Develop a set of indicators for estimating the cost of running the infrastructure and providing services. The indicators will have to be compatible and useable for all DiSSCo NTF members. Based on FR-NTF experience, an accurate yet simple method will be based on assessing the cost of the whole activity of DiSSCo member institutions. Therefore, conservation, access and infrastructure (e.g. equipment maintenance, building maintenance and administration) running costs will have to be accounted for to ensure DiSSCo activities can fully recover their costs. In order to allow comparisons across countries and institutions, the time effort of any staff time and the level of experience/salary of all staff will be incorporated. We will run workshops with the NTFs to set out and review the indicators, and review and incorporate previous work on staff costs/experience from EUColComp, SYNTHESYS 3 and ICEDIG. EuColComp is engaged in mapping all the activities occurring in any natural history museum and resulted in a competency*

*framework for those responsible for the collections. This set of competencies can be used in common job descriptions across European NSCs. During SYNTHESYS 3, a set of key performance indicators were developed to assess the activity in natural history museums: access and uses, digitisation, collections management and conservation, research and scientific activities, which can all be connected to resources. As a design study, WP3 in ICEDIG is currently surveying the digitisation methods and looking at the most efficient ones for 2D and 3D imaging.*

*Survey of costs and best practices for commissioning/decommissioning collections infrastructure. In addition to costs incurred in building IT infrastructures, we will survey the NTFs and other non-European museums to gather the cost of collection renovation projects in the last 10-15 years, as renovation projects aim, at least partly, at improving the conditions of physical access and therefore can be used by DiSSCo in setting up priorities, and possibly, including some of the costs in the service charges. We will include the costs of conservation and re-curation of collections, physical storage provision, and new buildings and supporting infrastructure. Any lessons learned and best practices will be summarised. Changes to collections infrastructure will be expressed as capital expenditure indicators in the cost book.*

*Costs of IT infrastructure and long-term preservation of data. This subtask will work closely together with WP6 to develop indicators on the running costs of the associated DiSSCo IT infrastructure both for the consortium and for individual institutions. This includes software tools and services, physical infrastructure and long-term data preservation.*

## **The DiSSCo Perimeter**

When it comes to cost calculation, it is important to identify the people who will be able to provide relevant information (mentioned as "contact persons" in the document) on the different activities. Such identification also helped to better define the DiSSCo perimeter. In parallel, there is a methodology provided by ESFRI on cost calculation for research infrastructures (Centre for Industrial Studies and European Strategy Forum on Research Infrastructures 2019). 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.

DiSSCo members (institutions) have activities which exist with or without the RI. These activities come at a cost and the RI will add some costs to act as a superstructure that allows a global offer to the users of the NSCs. 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.

There are two kinds of distributed infrastructure: those that bring different types of services under the same umbrella; and those where the services are common. DiSSCO falls into the latter category, as the services provided by the members are the same and the difference lies in the diversity of NSC's (taxon, origin, time...). The added value of DiSSCo is to be able to extract and to combine data from different sources and, therefore, the central hub requires a consistent and coherent knowledge of the cost of services provided by its members in order to interact efficiently with the users.

## **Natural Science Collections activities to be costed**

The Natural Science Collections-related activities identified during the project are listed below:

1. **Mass digitisation:** 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 5000 digitised specimens in a year (up to tens or hundreds of thousands in some cases). Digitisation does not automatically mean taking images; it may 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 5000 digitised specimens in a year. Digitisation does not automatically mean taking images; it may 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 quantitative and qualitative collection information which is not available through an externally accessible database: collections are not fully digitised. 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. The tooling-up of the scientific community in their different areas of expertise is also important. 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. Training may target both RI members and non-members.
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 for research purposes:** this covers time with decision-making committees for research loan requests, finding the requested specimens, assessing their condition, sampling fragments if requested, packaging and customs formalities. Time spent on loan requests for exhibition use is not included.

7. Physical access to collections for external researchers by visiting member institutions: the time accounted for starts with decision-making committees for these requests, to prepare the specimens for consultation, taking care of the material, organisation of the visit and dealing with the administrative requirement for hosting the visitor.
8. Asset maintenance: conservation and data curation. This covers the renovation of collections (relates to work sites, relocation of collections, restoration of a room etc.); preventative 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, including updates of digitised information).

### Illustrating the DiSSCo perimeter

The DiSSCo RI perimeter can be represented as concentric circles (Fig. 1). At the centre, there is the hub which coordinates the demand and the provision of services. Around this hub, the governing body is an ERIC (European Research Infrastructure Consortium) which can encompass other bodies than the hub: SLAs might be signed with institutions which would then become service providers on behalf of the ERIC. The outer circle is the Research Infrastructure: inside are the institutions which implement the eight categories of activity listed above. The addition of all the concentric circles represents the economic value and, at the same time, the economic impact of the RI. Fig. 2 demonstrates the activities carried out within the DiSSCo direct and indirect perimeters.

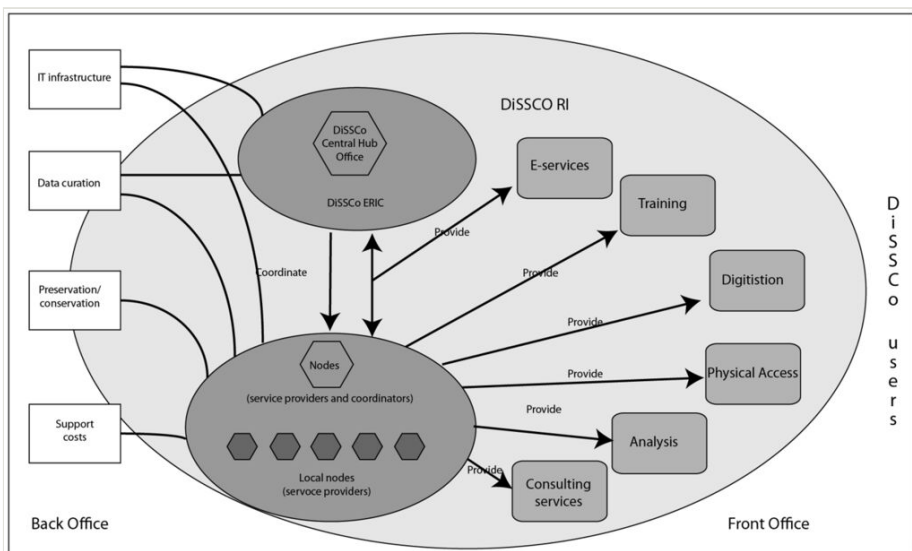


Figure 1. [doi](#)

Illustration of the DiSSCo perimeter.

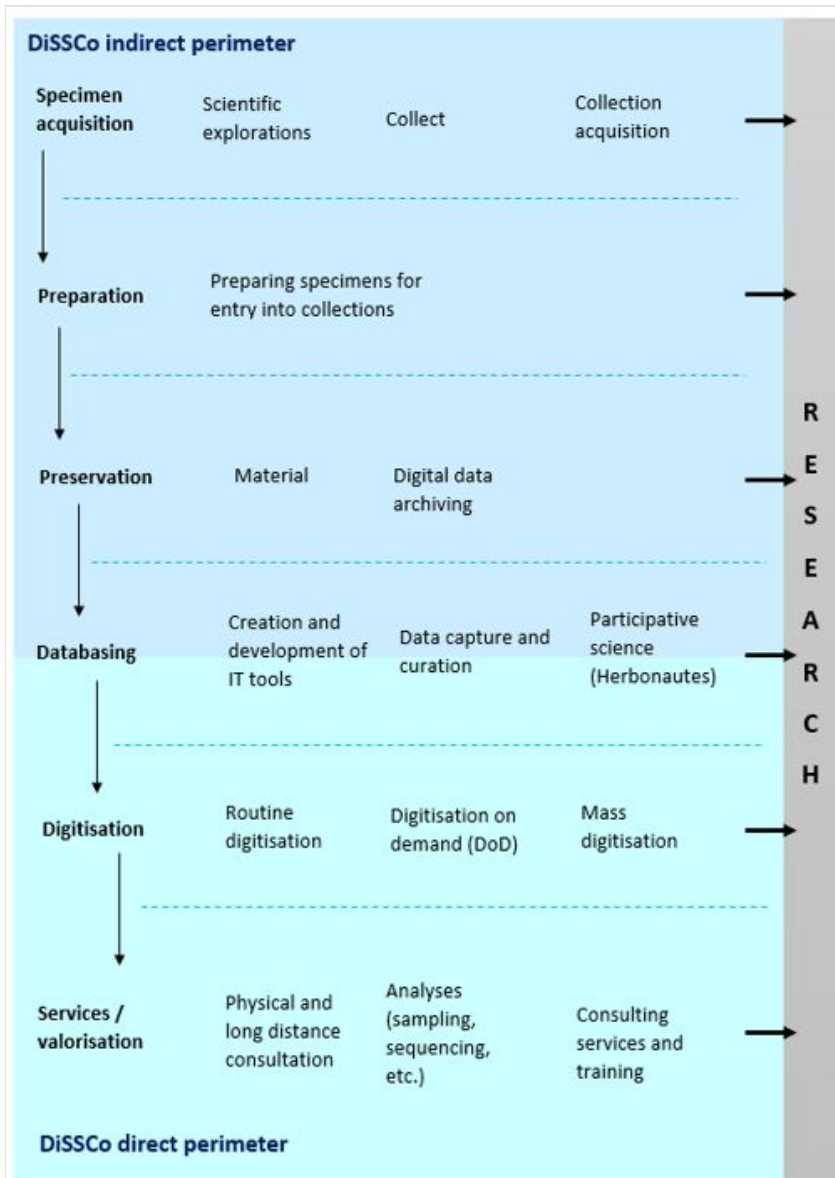


Figure 2. [doi](#)

DiSSCo RI perimeter at the Muséum national d'Histoire naturelle (MNHN, France).

## Previous experiences in RI cost estimation

Prior to DiSSCo Prepare, other projects took place at European and national level, two of which are relevant. 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. The National Network of Natural History Collections (RECOLNAT) French research infrastructure was one of the entities from which the costs were collected. Later, 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" (Hardisty et al. 2020b), 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. Research Infrastructures are often distributed over cross-cutting perimeters of actions distributed amongst different parcels of service or research units. It means that it is not possible to directly extract financial data from the different units.

### **ICEDIG: a first step towards the calculation of Natural Science Collections digitisation costs**

Between January 2018 and March 2020, a European Project, ICEDIG, aimed at preparing for the mass digitisation of European Natural Science Collections and the subsequent access to all related data in a harmonised and integrated manner across Europe. Three main deliverables from ICEDIG feed into the business framework of DiSSCo RI: the Cost Book of the digitisation infrastructure of DiSSCo (Hardisty et al. 2020a); Cost analysis of transcription systems (Walton et al. 2020a); and the Conceptual design blueprint for the DiSSCo digitisation infrastructure (Hardisty et al. 2020b).

### **Costs study: lack of information on digitisation and transcription systems costs**

The Cost Book of the digitisation infrastructure of DiSSCo (Hardisty et al. 2020a), is a study of 22 completed cost book templates from six institutions across Europe (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)). Following their experience in the digitisation of Natural Science Collections, 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 available for some collections (particularly herbaria where there has been longer-term and more widespread experience of digitisation) than others (e.g. vertebrates). Understanding that, the deliverable also explains that vertebrates and marine invertebrates are much more expensive to digitise than other collections (herbaria 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 two different steps:

1. establishing a digitisation facility, which is mainly capital costs; and
2. digitising specimens, which is mainly operating costs.

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

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. This study explains that data are not fully reliable, in particular "the lack of a common standards, data model and vocabularies have been a significant barrier to making these datasets comparable and interoperable".

The cost analysis of transcription systems (Walton et al. 2020a) 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 examining 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 report encourages institutions to quantify these costs in order to develop more efficient workflows.

### **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 (MESR) asked research infrastructures that are part of its RI 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.

### **General results**

Despite an excellent level of response, the results received were sometimes inconsistent and the degree of reliability of the results received was therefore partial. Not all

categories were always fully understood. Another fault of the methodology was that it was very time-consuming – which was partly due to the fact that, as it was new, there was an entry cost. In addition, the structure of the methodology changed during its implementation period. Some partners had already started working with the initial version. Regarding 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. One of the conclusions of the study was that annual dismantling costs, on average, never exceeded 5% of annual full operating cost. Some expenditure, such as construction and decommissioning, were not considered or were insufficiently calculated. 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 is likely to 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 1500 million euros. 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% (this probably understates actual cost). Capital costs are above the values that are usually calculated within RIs. Staff costs are below their usual weight within the budget, which is usually around 50%.

In terms of full-time equivalents (FTEs), in total around 7,000 FTEs participate in the operation of French and partly-French RIs. A small portion are directly employed by RIs. On average one FTE in the framework of French RIs costs annually 67000 euros. At international level, the annual cost per FTE is, on average, 136000 euros.

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 if the number of partners is high.

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.

## **Récolnat results**

The National Network of Natural History Collections (Récolnat) is a French research infrastructure (RI). It concerns all natural history collections and their valorisation by research. At the heart of its mission is the production and provision of a corpus of data for the study of current and past geology and biodiversity. At the time of the survey, the MNHN participated with the Récolnat research infrastructure. 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 science 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 32 million euros. The distribution of the costs was the following:

- 49% = amortisation;
- 8% = operating cost (except staff);
- 23% = staff costs;
- 20% = indirect cost.

## **Methodology implemented for the DiSSCo Cost Book**

Based on the perimeter of the RI presented above, two pilot projects were launched in February 2020. Based on those tests, contact 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, 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. Therefore, a methodology that was accessible for as many users as possible, despite language and cultural differences, had to be developed. 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 the implementation. It includes recommendations on how to encourage partners – notably those with no background in accounting or finance – to calculate the costs and outlines the financial rules selected for the exercise on NSC stakeholders.

## **Development of the Cost Book methodology**

### **The case of the Muséum national d'Histoire naturelle, Paris (MNHN)**

Once the MNHN perimeter was delimited, a pilot project was implemented between October 2020 and January 2021 in order to test an initial methodology. It was mainly inspired by the cost methodology proposed by the MESR in 2016 and concerned both 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 main assumption was that staff cost would represent the bulk of the total cost and, therefore, the focus should be to distribute staff time across the different activities of the RI perimeter. As this information is not necessarily recorded even if the organisation of the department matches the structure of the RI, the methodology allows for uncertainty. These uncertainties were then used in simple error propagation tools.

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.

As there was no unified framework, it was complex to define a harmonised methodology to trace the costs and gathering useful data, all of which relied on good interaction with the managers leading the different departments.

Thanks to this first test, MNHN identified the information that was essential and non-essential. 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!

### **Key results from MNHN**

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

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

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 in Fig. 3 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.

The MNHN is a large institution which is structured via collection departments corresponding to the different natural history disciplines (botany, zoology etc.) and different analytical platforms (CT-scanning etc.). 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 (Fig. 4). The method to assess these indirect costs could not be implemented at the MNHN for lack of relevant data.

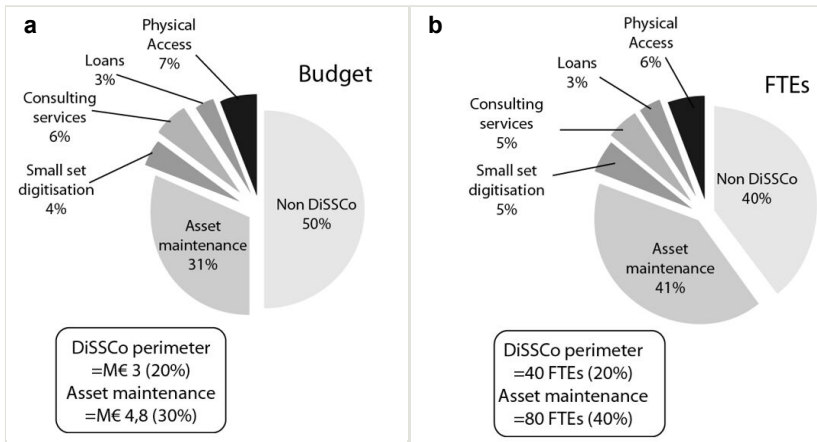
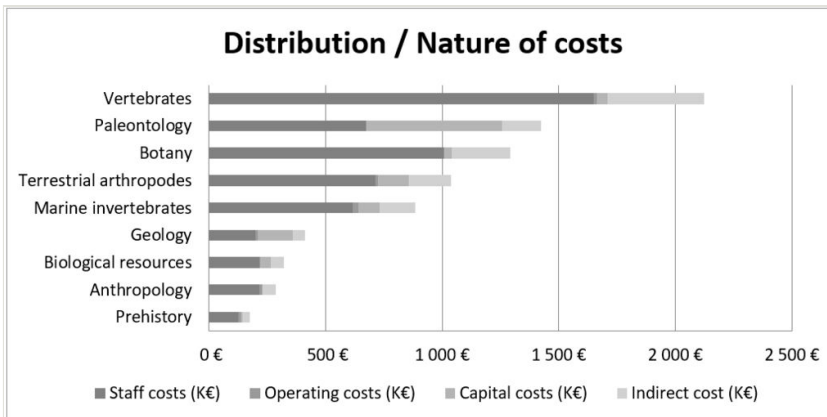


Figure 3.

MNHN Collection department budget and FTEs distribution (2019 data):

**a:** MNHN Collection department budget (2019 data); [doi](#)

**b:** MNHN Collection department FTEs distribution (2019 data). [doi](#)

Figure 4. [doi](#)

MNHN collection department detailed structure of costs (2019 data).

## 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.

Fig. 5

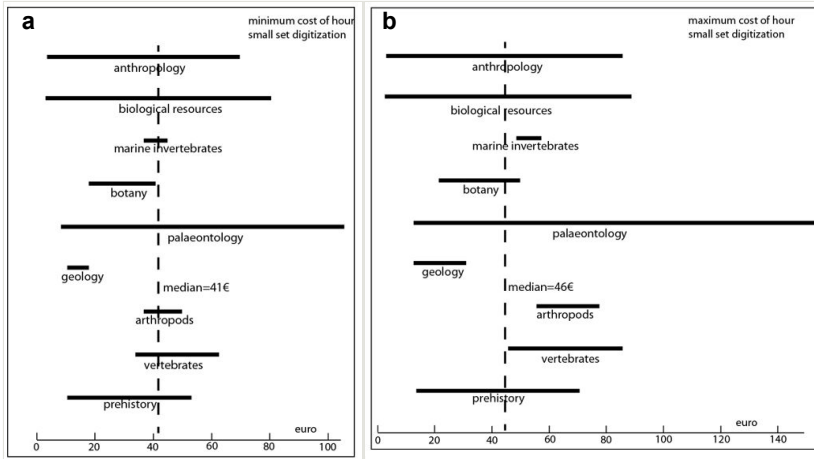


Figure 5.

Cost per hour of small set digitisation at MNHN – Minimum cost and maximum cost of an hour for small set digitisation shown with calculated error bars (2019 data):

**a:** Minimum cost of an hour for small set digitisation at MNHN shown with calculated error bars (2019 data); [doi](#)

**b:** Maximum cost of an hour for small set digitisation at MNHN shown with calculated error bars (2019 data). [doi](#)

Fig. 6

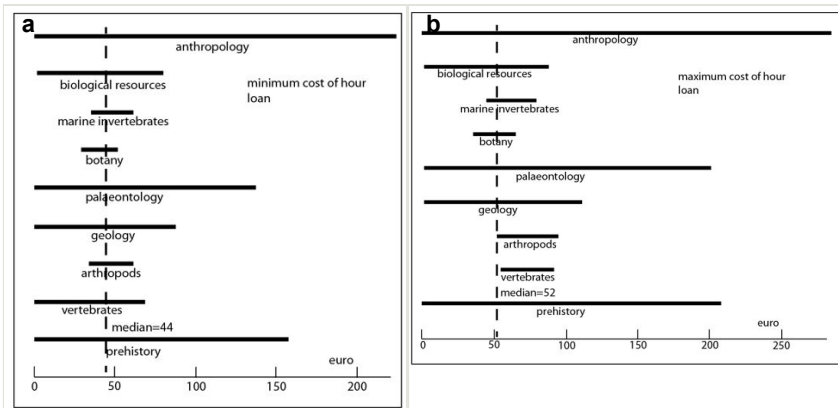


Figure 6.

Minimum and maximum cost per hour for loan at MNHN, shown with calculated error bars (2019 data):

**a:** Minimum cost per hour for loan at MNHN, shown with calculated error bars (2019 data); [doi](#)

**b:** Maximum cost per hour for loan at MNHN, shown with calculated error bars (2019 data). [doi](#)

Fig. 7

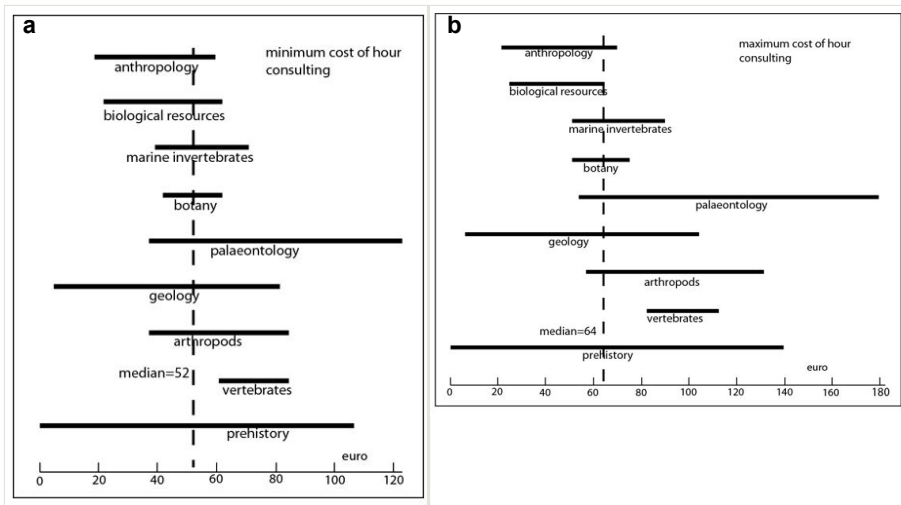


Figure 7.

Minimum cost and maximum cost of an hour for consulting at MNHN, shown with calculated error bars (2019 data):

**a:** Minimum cost of an hour for consulting at MNHN shown with calculated error bars (2019 data); [doi](#)

**b:** Maximum cost of an hour for consulting at MNHN shown with calculated error bars (2019 data). [doi](#)

Comparing the cost/hour for the different services and collection department (Fig. 5, Fig. 6 and Fig. 7) leads to two conclusions:

- The uncertainty is variable for each department; this is related to the number of staff in the department: the larger the number, the more precise the estimation. The large number is not only influenced by the number of staff, 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 is 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.

## Analysis of MNHN cost per service (digitisation, loans, consulting and visits)

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 (Fig. 8): 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 levels of data collection adopted during digitisation.

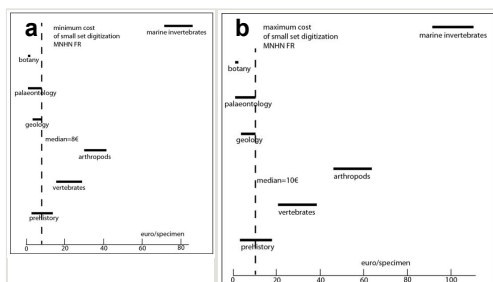


Figure 8.

Minimum and maximum cost per digitised specimen for small set digitisation at MNHN (2019 data):

**a:** Minimum cost per digitised specimen for small set digitisation at MNHN (2019 data); [doi](#)

**b:** Maximum cost per digitised specimen for small set digitisation at MNHN (2019 data). [doi](#)

Loans (Fig. 9): low value for geology and high value for botany would require examining the procedures used by these two services. There are likely 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).

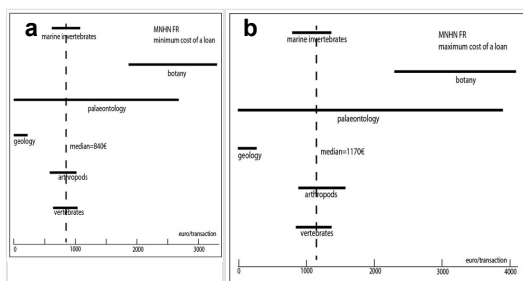


Figure 9.

Minimum and maximum cost per loan at MNHN (2019 data):

**a:** Minimum cost per loan at MNHN (2019 data); [doi](#)

**b:** Maximum cost per loan at MNHN (2019 data). [doi](#)



Visits (Fig. 10): 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.

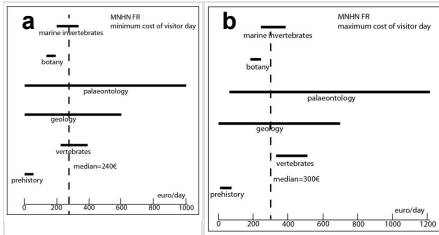


Figure 10.

Minimum and maximum cost per physical visit at MNHN (2019 data):

- a: Minimum cost per physical visit at MNHN (2019 data); [doi](#)
- b: Maximum cost per physical visit at MNHN (2019 data). [doi](#)

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.

MNHN Collections department productivity data (Fig. 11): for any project the question of its duration is as important as the question of its cost. It is relevant to estimate services with time units as well as or instead of currency.

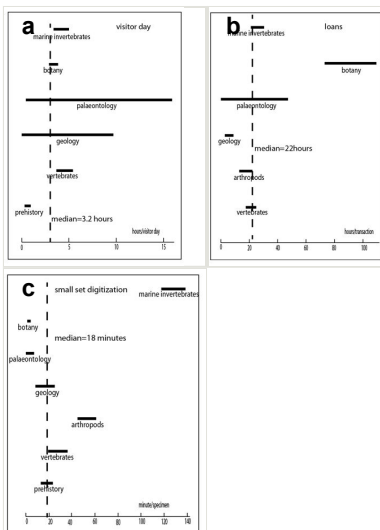


Figure 11.

Comparison of cost areas and hours per service at MNHN (2019 data):

- a: Comparison of cost areas and hours for visitor day at MNHN (2019 data); [doi](#)
- b: Comparison of cost areas and hours for loans at MNHN (2019 data); [doi](#)
- c: Comparison of cost areas and hours for small set digitization at MNHN (2019 data). [doi](#)

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 mission and aims. Digitisation of collections is a good option as it will save time; it will help in preparing access or implementing remote working. A digitisation programme will save resources in the institutions, those resources may then be deployed to document and provide better support to visitors, who will have better prepared their visits, providing both better efficiency and better effectiveness overall.

Table 1

Table 1. Overview - cost calculated based on MNHN services (2019 data).				
EURO	Consulting services	Physical access	Loans	Small set digitisation
Minimum cost		1101€ per visit	839€ per loan	8€ per specimen
Maximum cost		1472€ per visit	1171€ per loan	11€ per specimen
Minimum cost per hour	52€	47€	44€	41€
Maximum cost per hour	64€	59€	52€	46€

MNHN analytical facilities results (Fig. 12 and Fig. 13).

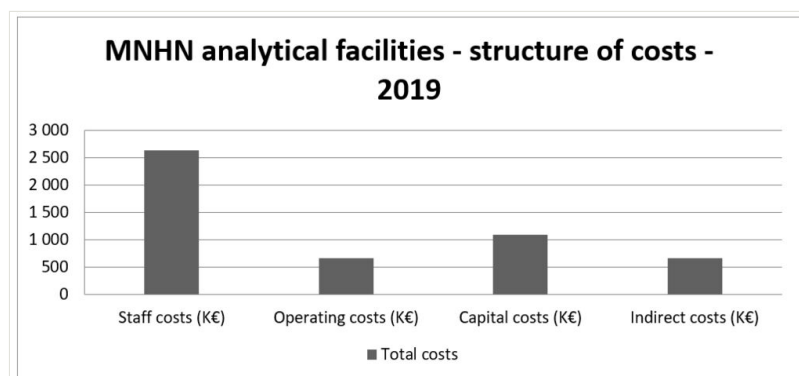


Figure 12. [doi](#)

MNHN analytical facilities - structure of costs (2019 data).

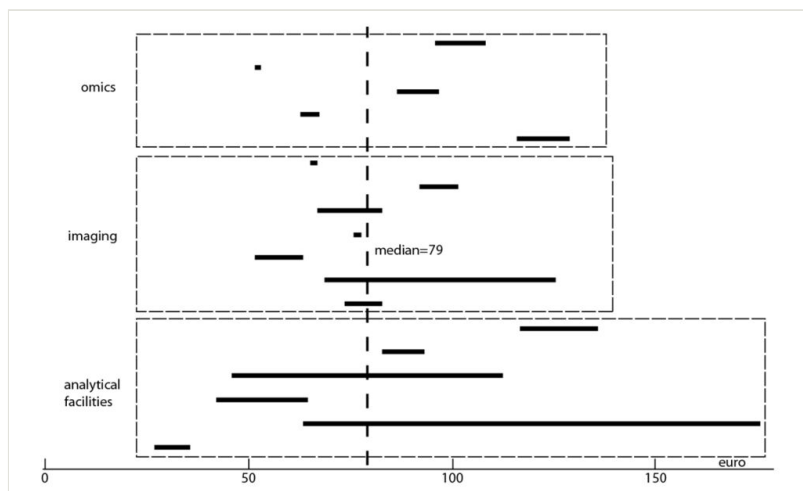


Figure 13. [doi](#)

Cost per hour of analytical facilities at MNHN (2019 data).

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

## Tests at European level

In order to test the methodology with DiSSCo stakeholders speaking different languages, having different accounting systems and different work practices, a new version of the methodology was shared with project partners (NHM, RBINS, APM, SGN). 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 discuss 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. 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 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.

## General methodology requirements

The following elements guided the implementation of the methodology with a large number of partners not necessarily fully aware of the DiSSCo process.

- Identify contact persons with a good understanding of staff activities, annual operating and investment costs. They are pivotal in the Cost Book process; generally, they are the managers and senior staff.
- Explain the objective to bring along the contact persons. There were two objectives: (i) provide an estimate of the overall contribution of the institutions to DiSSCo RI as an element of the political discussion with governments; and (ii) estimate the costs paid to provide services. This would allow partners to have a better understanding of their own costs and, if needed, be able to charge for some of their services.
- Focus on what matters. Accounting and reporting structures are diverse within 170 different institutions across 23 countries and do not necessarily match the structure in the methodology. Therefore, 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, to minimise the effort by the contact person. However, in order to get the same data from all partners, the structure of data should not be changed over the course of implementation as that would increase complexity and decrease consistency in processing results.
- Find the right granularity, using the concept of accounting entities as defined in the Financial Glossary: "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. Having the right accounting entity for which costs could be mastered was considered a better option than imposing categories such as botany, vertebrates etc., in order to give institutions the possibility to apply the methodology to their own structure and organisation. The contact persons can then be the heads of the different accounting entities of the organisation.
- Communicate and keep in contact. The methodology should be short, without too much text and, if possible, with images and diagrams. This allows for a rapid understanding of the concepts used and, in case there is a question, people can contact the team. However, people tend to not read documents or even emails and, therefore, it is necessary to regularly send reminders. As a consequence, it is important to have reasonable deadlines.
- Account for uncertainty. Due to the lack of similarity between accounting and reporting systems and the methodology, data gathered may not be accurate. Introducing four levels of uncertainty (from exact value to poor knowledge) 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 for costs.

## Costs

Direct costs correspond to staff costs, operating costs and investments directly connected to the project. The conclusions of the two pilot projects are that staff costs are the main costs in the DiSSCo context. These costs require interviewing managers on the distribution of time in their teams, access to the salary levels of all the people involved in services within the DiSSCo perimeter and whether staff are on part-time or full-time employment.

Operating costs other than staff costs have a minor impact on the total costs. Therefore, although they are not always clearly identified, it is possible to use estimations and a rough distribution within the different DiSSCo cost areas.

Investments/capital costs are a bit more complex to evaluate. According to the ESFRI methodology, amortisation cannot be considered in the cost calculation process. This is relevant for the hub and the IT infrastructure which constitute the centralised costs of the RI which is being built. It is not relevant for the institution's contribution for the purpose of this methodology which aims at estimating the full costs of the services provided by the institutions. Due to the lack of information regarding past investments and the irrelevance of using annual budgets for a multi-annual investment, it was decided to use a yearly average of the investments of the 5 years preceding the cost assessment.

Indirect costs are the support costs covered by the institution that are spread over 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).

Ideally, these indirect costs could be estimated if the data which are required, such as electricity consumption per service, are available. Unfortunately for most institutions, including MNHN, this is not yet possible and, therefore, the decision was taken to use the flat rate of 25% on direct costs as proposed by the EU. However, institutions using the proposed methodology have the possibility to use their own flat rate, based on their own calculation of their indirect costs if they are able to do so.

## The package 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. Suppl. material 1 shows snapshots of the Excel tool that was provided to member institutions and national nodes.

### **General information with automatic impact (see Supplementary material 1)**

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 post (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 (Suppl. material 1).

### **Distribution of staff time (see Supplementary material 2)**

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 (Suppl. material 2).

### **Operating and capital costs (see Glossary)**

The next two steps are related to operating costs and capital costs (see Financial Glossary for definition). From the MNHN pilot project, staff costs were assumed to be significantly larger than other costs and, therefore, only rough estimations were required. In this respect, partners were asked to add a general annual cost for both categories and to distribute them in percentage terms amongst the exact same categories as for the staff costs.

### **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 that accounting entity;
- Proportion of staff costs in relation to total costs;
- The total costs of each DiSSCo area implemented within the accounting entity;
- 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.

From these results, the team produced a set of recommendations (Suppl. material 3).

### **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 amongst 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 provided; a cost per loan; a cost per specimen loaned; a cost per visitor; and a cost per visitor day. This allows for a comparative approach amongst DiSSCo member institutions.

For each cost per category, there is a minimum and 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 total costs of the accounting entity. 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 current related services. With these data, it is possible to estimate the number of FTEs involved in the work on Natural Science Collections at European level and, therefore, what kind of costs are linked to the collections.

## **Methodology for centralised costs: Central Hub Office and IT services**

Contrary to the cost of services in the national nodes and institutions, the cost of the Central Hub is the cost of the superstructure piloting the RI. Therefore, the method was more akin to standard cost estimation of any administrative structure. The difficulty resided in defining the different roles needed to run an infrastructure like DiSSCo.

However, as for the methodology for the RI members, 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. It is also necessary to take account of inflation and of the country where the DiSSCo headquarters may be situated. Finally, indirect costs are calculated, based on the EU rules: 25% of the direct costs (excludes subcontracting).

### **Central Hub office cost calculation**

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 for performance and compliance (including standards and policies).
6. Representation: EU/international relations (incl. expansion).
7. Core service provision: coordination of provision of services, standards and policies.
8. User support: immediate support at institutional level.
9. Communication and outreach: public relations, communication.

### **IT infrastructure cost calculation = costs to operate the Central Hub digital tools and services**

The needs for DiSSCo IT infrastructure are addressed elsewhere (Islam et al. 2020, Islam et al. 2022, Islam et al. 2020, Leeflang et al. 2022). What we can say here is that some associated costs are highly variable according to the data storage model of the RI. Either the data are archived by the central hub or they are archived by the member institutions.



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, but the bulk of the storage costs will be incurred by the member institutions. Therefore, although the metadata preserved by the hub will grow over time, the resources required for preserving the data of the RI itself will be relatively small. There are specific issues to deal with regarding the DiSSCo infrastructure:

- Timeline: costs must account for the separate construction and operation phases. The construction phase can take several years and, therefore, the costs of construction are spread over this number of years;
- Staff cost: FTEs and average cost per FTE will depend on the location of the host country;
- Outsourcing cost: subcontractor/annual licences;
- Cloud computing/Hosting cost: storage (short and long-term);
- Physical assets and decommissioning costs;
- Possible contribution by member institutions to the different components of the IT architecture;
- User thresholds above which the components of the architecture should be redesigned (scaling of the infrastructure);
- Level of service provided by the IT infrastructure;
- Training to engage users;
- Translation from operation costs to major upgrades.

## Results

### 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 users do not always read the whole guide and instead skip to directly working 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 fully reflect 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 data are copied and pasted, for example, the formulae can disappear. A possibility is to harmonise training 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.

## Cost estimations

### Central Hub

A first estimation is that the construction phase would cost around 2.5 million euros over a period of two years. DiSSCo operation would annually cost 2.6 million euros. The costs to decommission the RI would be between 2 to 4 million euros (1 to 2 years of operation).

Given the function of the central hub office, there is no significant difference between the construction phase and the operational phase (respectively 1.4 M€/year and 1.42 M€/year). The costs of the DiSSCo Central Hub Office will be mainly staff costs (62%) compared to operating costs (17%) and indirect costs (20%). Capital costs are low as DiSSCo does not require the construction of a physical infrastructure.

The total cost to build DiSSCo IT services and architecture is estimated at 2.2 million euros; the annual cost would vary depending on the length of the construction phase (2 or 3 years). The cost to annually run and maintain the DiSSCo IT services and architecture during the operation phase is estimated at 1.2 million euros.

Similarly to the central office, the nature of the costs is mainly staff costs; they represent around 60% in the construction phase and 50% the operational phase. Direct operating costs (outsourcing, hosting, upgrades) represent around 25% of the construction phase and 35% of the operation phase.

### DiSSCo national nodes, cost per hour

Twenty-seven institutions from nine countries (AT, BE, DE, DK, EE, FR, IT, NL, UK) answered the Cost Book survey (including MNHN results), totalling information on 76 accounting entities. 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 (Fig. 14, Fig. 15 and Fig. 16, Fig. 17).

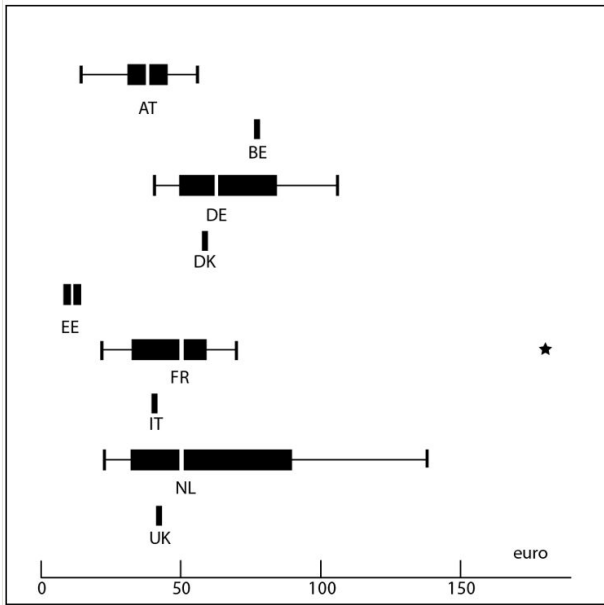


Figure 14. [doi](#)  
 Minimum cost per hour for small set digitisation for the national nodes.

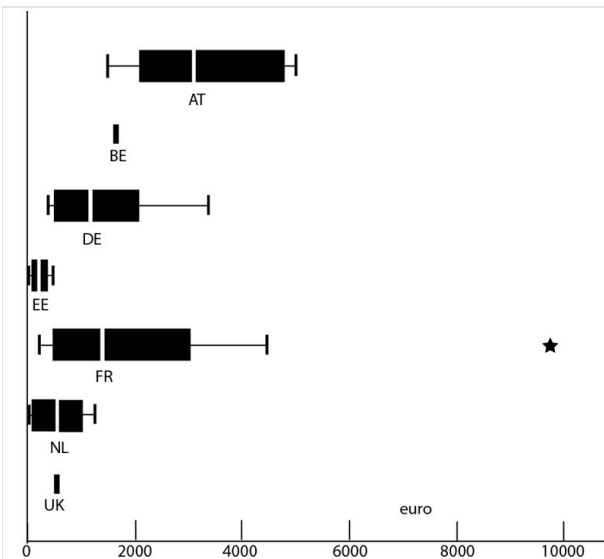


Figure 15. [doi](#)  
 Minimum cost per loan for the national nodes.

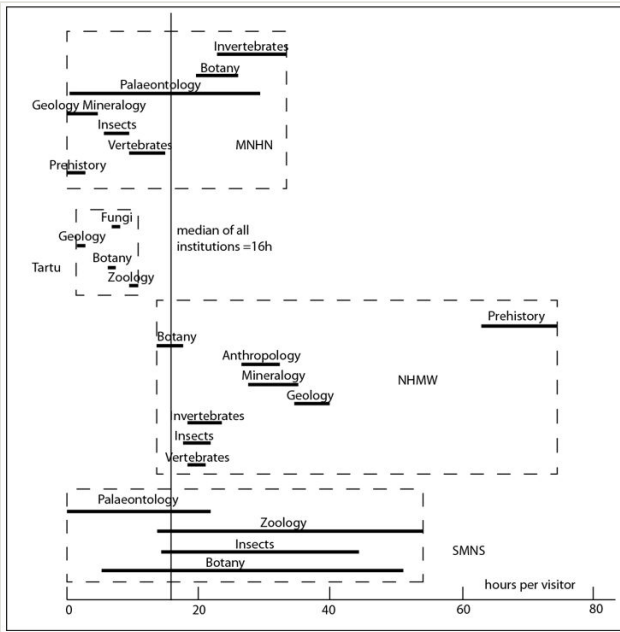


Figure 16. [doi](#)  
 Number of hours per visitor for each Accounting Entity (AE) at the institution.

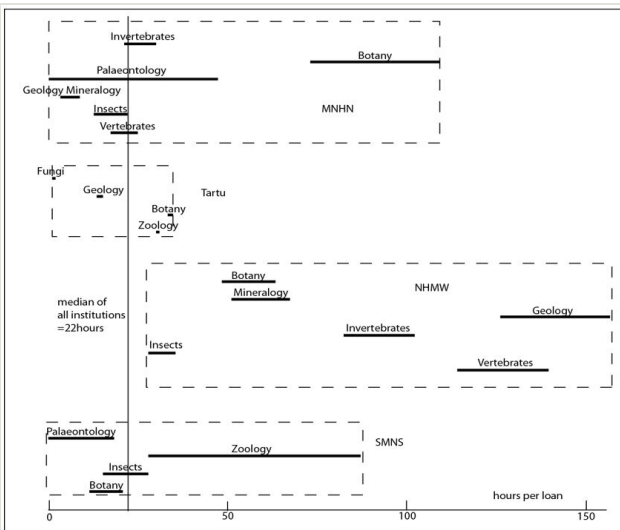


Figure 17. [doi](#)  
 Number of hours per loan for each AE in the institution.

Even if the outlier in the FR node is taken out, the spread of the data remains quite large, with minimum-maximum values ranging between 10-100 euros.

As for the cost-per-hour for small set digitisation, there is one outlier in FR node, but removing it from the dataset still leaves a significant spread of data.

Further work is needed to check whether this discrepancy originates in structural differences reflecting 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.

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

This figure shows 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 amongst all the institutions.

The figure for loans (Fig. 17) shows the same pattern as hours per visitor (Fig. 16). As for the cost data (Fig. 14, Fig. 15), the median is the best indicator of the distribution of the data.

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 accounting entities. Understaffing of some collection departments and high share of equipment cost in the 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 2

Table 2. Median values for the different services of all responding institutions.							
€/hour	Mass digitisation	Small set digitisation	Consulting services	Training	Analysis	Loans	Physical access
Minimum	45	48	55	56	65	44	48
Maximum	56	58	62	73	82	57	60

Consulting services, training and analysis generally require more seniority and specialist skills than digitisation, loans and physical access, a difference which will be weighed in the cost per hour of the staff.

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 both surveys.

Table 3

Table 3. Median values for the different services.			
	Unit	Minimum cost	Maximum cost
<b>Mass digitisation</b>	cost per specimen	2.13 €	2.74 €
<b>Small set digitisation</b>	cost per specimen	13.10 €	15.42 €
<b>Loans</b>	cost per loan	1 043 €	1 267 €
<b>Physical access</b>	cost per visitor	707 €	858 €
	cost per visitor day	204 €	252 €

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

### Conclusions on the results of the consultation survey

The distribution of 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 accounting entities 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 (digitisation associated with loans, different MIDS level of digitization etc.)

### 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 (out of 170). Besides, as shown above, the diversity of the institutions and the way they understand the Cost Book makes the extrapolation to 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. With this due caution, the services provided by the institution's members of DiSSCo would be around 290 million euros 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 4

Table 4. Calculation of the DiSSCo costs declared.	
	<b>Total annual costs declared €</b>
Mass digitisation	4,899,504
Small set digitisation	3,942,355
Consulting services	3,230,765
Training	1,284,268
Analysis	9,237,297
Loans	2,392,562
Physical access	3,082,940
Sub total	46,257,979
Asset maintenance	18,188,287
Total costs DiSSCo	75,663,706
Total non-DiSSCo costs	29,405,726

The set of institutions that have answered the survey may not be representative of the DiSSCo members; therefore, extrapolating from these 27 answers to the whole DiSSCo RI is probably biased. However, it gives an order of magnitude of what the DiSSCo members are already spending on the services. Assuming that the sub total of 46,257,979 euros represents 27/170 of the total cost of DiSSCo services, the overall cost of DiSSCo services amounts to an estimated 291,253,942 euros.

## Conclusion

The DiSSCo Cost Book aimed at calculating the costs to build, operate and dismantle the DiSSCo Research Infrastructure. To do so, we needed to understand 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 characterises a distributed research infrastructure like DiSSCo is its decentralisation, its potential for evolution according to funding opportunities and the willingness of its members to increase their participation in the RI. During the early stages, there is a lot of scope for future strategic development.

The 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 partially and separately by DiSSCo member institutions who preserve Natural Science Collections, share data and provide services. The goal was to find a way to encourage them to quantify the costs associated with those activities. To do so, we developed a methodology accessible to as many people as possible and sharable despite 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 costs shared at European level.

The results of this strategy are partial: 27 institutions out of 170 targeted shared data. Nevertheless, the main results are that, in proportion to the results received, annually 290 million euros 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-service, are in comparable price categories. The methodology developed can, therefore, 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 decentralised costs.

Concerning the centralised costs, the 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 1.4 million euros. This cost can change over time. Often, research infrastructure actions grow and initial budgets are lower. 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 2.2 million euros in total (excluding the cost of the Central Hub office). The annual cost to maintain and run DiSSCo IT infrastructure could be around 1.2 million euros. 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. In addition, perhaps most importantly, it defines cost units, providing a framework that can be used for future cost calculation exercises in the field of Natural Science collections.



## DiSSCo Financial Glossary

Table 5

Table 5. DiSSCo Financial Glossary	
Accounting entity	An entity based on the principle of subsidiarity (a principle of social organisation 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.
Amortisation of DiSSCo data	Amortisation 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	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, for example, 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	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 costs and capital costs 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 an employee to their employer. Considers all charges associated with employees (includes sums paid under the employee savings or pension schemes, 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 1607 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: 1. 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; 2. 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 modernisation 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 non-durable goods. 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 non-durable goods. Example: an annual contract for maintenance worth 10,000 euros is an operating cost as it is not a durable good. Example 2: Purchase of a pack of notebooks worth 50 euros = operating cost.
Staff costs	Expenditure incurred for staff time used to deliver projects, for example, annual salary, national insurance, pension contributions, employer's contributions for NI and pension, any other contractual payments included in the employee contract.

## DiSSCo Perimeter Glossary

Table 6

Table 6. 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 preventative 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, including 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 digitised, 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	Corresponds 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 data standards and levels collected for different purposes. This cost area includes staffing costs associated with improving data curation activities (e.g. time on external projects developing or improving standards and practices such as participation in the collections descriptions interest group; or technical staff time on internal data curation or audit projects).
Digitisation	Two possibilities: mass digitisation and routine digitisation. Digitisation includes the process of converting analogue 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 by 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	<p>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 5000 digitised specimens in a year (up to tens or hundreds of thousands in some cases). Digitisation does not automatically mean imaging, it can also mean only databasing.</p> <p>By mass digitisation, we mean digitising 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).</p>

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 each 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, organisation of the visit and dealing 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.
Preventative 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.
Running IT infrastructure	<p>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:</p> <p>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.</p> <p>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.</p> <p>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.</p>
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 5000 digitised specimens in a year. Digitisation does not automatically mean imaging, it can also mean only 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 (Castelin et al. 2020).
----------	---

## Funding program

[H2020-EU.1.4. - EXCELLENT SCIENCE - Research Infrastructures](#) Main Programme

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

## Grant title

DiSSCo Prepare – "Distributed System of Scientific Collections - Preparatory Phase Project", Grant Agreement No. 871043

## Author contributions

**Salomé Landel:** Writing - original draft (lead); visualisation (lead); conceptualisation (equal); methodology (equal). **Michel Guiraud:** Methodology (lead); supervision (lead); writing – review and editing (supporting). **Ana Casino:** Project administration (supporting); methodology (supporting); supervision (supporting); writing – review and editing (supporting).

**Lisa French:** Writing – review and editing (supporting); conceptualisation (supporting); methodology (supporting). **Laurence Livermore:** investigation (supporting); resources (supporting). **Helen Hardy:** Methodology (supporting); investigation (supporting); Writing – review and editing (supporting). **Eva Alonso:** Methodology (supporting); project administration (lead). **Wouter Addink:** Methodology (supporting); supervision (supporting); investigation (supporting); project administration (supporting). **Sharif Islam:** Methodology (supporting); supervision (supporting); investigation (supporting); project administration (supporting). **Dimitris Koureas:** Methodology (supporting); project administration (lead); supervision (supporting). **Katharine Worley:** Writing – review and editing (supporting). **François Dusoulier:** Writing – review and editing (supporting); conceptualisation (supporting); methodology (supporting). **Marko Lovrić:** Writing – review and editing (supporting); investigation (supporting). **Claus Weiland:** Writing – review and editing (supporting); investigation (supporting). **Hanieh Saedi:** Writing – review and editing (supporting); investigation (supporting). **Stefaan Pijls:** Writing – review and editing (supporting); investigation (supporting). **Frederik Leliaert:** Writing – review and editing (supporting); investigation (supporting). **Patricia Mergen:** Writing – review and editing (supporting); investigation (supporting). **Carole Paleco:** Writing – review and editing (supporting); investigation (supporting). **Serge Scory:** Writing – review and

editing (supporting); investigation (supporting). **Patrick Semal**: Writing – review and editing (supporting); investigation (supporting). **Gaël Lymer**: Writing – review and editing (supporting); investigation (supporting).

## Conflicts of interest

The authors have declared that no competing interests exist.

## References

- Castelin M, Blettery J, Paleco C, Rey I, Santos C, Babocsay G, Mergen P, Giere P, Akkari N, Schiller E, Schweiger S, Wiltshcke K, Holtstam D, Langhof J, Fulcher T, Paton A, Smirnova L, Theeten F, Aronsson H, Obst M, Boer Hd, Steenberge MV, Casino A, Tilley L (2020) D2.3 Catalogue and recommendations for the development of a proactive, efficient and evolving DiSSCo training programme. URL: <https://know.dissco.eu/handle/item/349>
- Centre for Industrial Studies, European Strategy Forum on Research Infrastructures (2019) Guidelines on Cost Estimation of Research Infrastructures. STR-ESFRI. URL: <https://www.esfri.eu/latest-esfri-news/new-study-guidelines-cost-estimation-research-infrastructures-str-esfri>
- Cigref (2018) Modèle d'analyse et de benchmarking des coûts informatiques. Cigref. 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](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)
- Hardisty A, Livermore L, Walton S, Woodburn M, Hardy H (2020a) Costbook of the digitisation infrastructure of DiSSCo. Research Ideas and Outcomes 6 <https://doi.org/10.3897/rio.6.e58915>
- Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalgo A, Paul D, Runnel V, Vermeersch X, van Walsum M, Willemse L (2020b) Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1. Research Ideas and Outcomes 6 <https://doi.org/10.3897/rio.6.e54280>
- Islam S, Hardisty A, Addink W, Weiland C, Glöckler F (2020) Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections. CODATA Data Science Journal 19: 50. <https://doi.org/10.5334/dsj-2020-050>
- 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 <https://doi.org/10.3897/rio.8.e93816>
- Landel S, Guiraud M, Casino A (2023) DiSSCo Prepare Deliverable D4.1 - The Cost Book for DiSSCo. DiSSCo Prepare. <https://doi.org/10.34960/kdkr-sf06>
- Leeflang S, Weiland C, Grieb J, Dillen M, Islam S, Fichtmueller D, Addink W, Haston E (2022) DiSSCo Prepare D6.2 Implementation and construction plan of the DiSSCo core architecture. DiSSCo Prepare. <https://doi.org/10.34960/50b9-kj05>

- 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 <https://doi.org/10.3897/rio.6.e56211>

## Supplementary materials

### Suppl. material 1: Appendix 1 [doi](#)

**Authors:** Salomé Landel, Michel Guiraud

**Data type:** Cost assessment data

**Brief description:** Snapshots of the cost estimation tool developed for the DiSSCo Cost Book.

[Download file](#) (276.77 kb)

### Suppl. material 2: Appendix 2 [doi](#)

**Authors:** Salomé Landel, Michel Guiraud

**Data type:** List

**Brief description:** List of institutions that participated in the DiSSCo Cost Book study.

[Download file](#) (100.34 kb)

### Suppl. material 3: Appendix 3 [doi](#)

**Authors:** Salomé Landel, Michel Guiraud

**Data type:** Set of recommendations

**Brief description:** DiSSCo Cost Book Recommendations.

[Download file](#) (86.19 kb)