



# INFRA-FED 2022

## Impulse action call

# PROPOSAL DESCRIPTION

Please consult the Information File before filling out this template. This document must be submitted in **WORD and PDF** form via email to [infra-fed@belspo.be](mailto:infra-fed@belspo.be), naming the file **Proposal Description\_[project acronym]**. The subject of the email must be: **INFRA-FED 2022 FullProp\_[project acronym]**. Please respect the font type and size given by the template.

The deadline for the submission is: **Tuesday 6<sup>th</sup> December 2022 @ 14h00**.

TODO Risk assessment, Gant chart, DMP, WP description

## PART A: GENERAL INFORMATION OF THE PROPOSAL

### 1. SUMMARY

#### CANATHIST

**The Natural History collections collected in central Africa by Belgian institutions**

**3 years | 725631,00 €**

Collections of natural history reflect both the unique worldwide biodiversity of past and present life and the geological history of the Earth. It is estimated that European Natural history (NH) collections contain up to 1 billion specimens. When used at their full potential, NH collections are able to contribute to understanding and mitigation of critical challenges such as biodiversity loss, climate change, invasive species or the vectors of diseases. The collections of the Africa Museum (RMCA), the Royal Belgian Institute of Natural Sciences (RBINS) and the federal collection of Meise Botanic Garden (MBG) form together the richest collection (estimation of about [10 million specimens](#) and more than [100000 type specimens](#)) related to Central Africa due to the colonial history of Belgium. The collections of the RMCA and RBINS are moreover concerned by the law ([loi du 3 juillet 2022](#)) concerning the process of restitution of colonial collections and the draft bilateral agreement between Belgium and the Democratic Republic of Congo. The provenance study and the digitization of these collections is therefore a priority for this unique heritage and will contribute to the Belgian federal input in the setup of the [DiSSCo](#) European research infrastructure. A significant part of the FSIs natural history collections related to Central Africa have been made during the colonial period in the framework of scientific expeditions (e.g. Expeditions of the National Parks of Congo). The results of these field works were often divided between RMCA, RBINS and MBG. After the independence of the Congo, a part of the collections were also prepared for repatriation to Congo and stored in the Africa Museum, still waiting for a formal agreement with a host institution in DRC.

The goal of this proposal is to produce the most complete inventory of the concerned collections among the Belgian DiSSCo institutions and non-DiSSCo scientific institutions as research infrastructure, make a provenance study, prioritize the digitization of the collections (data, metadata and digital specimens) and share all information, digital specimens and archives with the competent authorities, institutions and citizen communities of DRC, Burundi and Rwanda, using a specific collaborative infrastructure and/or dedicated offline tools. There is a need, not only to give access or provide to overseas countries the FAIR access to the digital collections related to their own country but also to offer to them, as well as to the FSIs and other users/clients of Belnet a set of integrated open source tools (Collection Management systems, 2D Images server, 3d object server, Map server and Annotation server (text and image) ) allowing them to contribute to international infrastructures within the setup of EOSC.

This digital collection will also contribute to the ESFRI DiSSCo infrastructure with unique data which are today strongly under-represented in the world's biodiversity inventory. This support to DiSSCo will help the 2 Federal scientific Institutions to participate in this community. Finally, the developed technical infrastructure will offer a generic and powerful set of tools to the other members of the Belgian DiSSCo communities and to the other FSIs allowing them to contribute to the different international networks/infrastructures related to the cultural and scientific heritages.

CANATHIST aims to reuse, optimize and share tools developed during previous Open Source projects (Agora 3D, DIGIT, NaturalHeritage.be) or existing already in the Belnet portfolio and will include a strong component of capacity building and participation of non-funded partners in the integration of the mentioned tools through specific training and technology transfer.

*1; DiSSCo, 2; Provenance study, 3; Digitisation, 4; Restitution, 5; Central Africa, 6; Open data.*

## PART B: PROMOTOR / PARTNERSHIP

### COORDINATOR

#### COORDINATOR (P1): RMCA

Civility	<input type="checkbox"/> Prof.	<input checked="" type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Didier		
Last Name	Van den Spiegel		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Royal Museum for Central Africa		
Department	Zoology		
Service	Collection and data management		
Email	didier.van.den.spiegel@africamuseum.be		
Telephone	+32496444904		
Website	www.africamuseum.be		
Full name of Director General (DG) / President of the Institution	Karel Velle		
Official function of DG / President	General Director <i>a.i.</i>		
Email of DG / President	dir@africamuseum.be		

### OTHER PROMOTORS FINANCED BY THE PROJECT

Two tables are provided here, if there are more promoters financed by the project, please replicate the table as many times as necessary, adding P4, P5...

#### P2: [RBINS]

Civility	<input type="checkbox"/> Prof.	<input checked="" type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Patrick		
Last Name	Semal		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Royal belgian Institute of natural Sciences		
Department	Scientific service of Heritage		
Service	Anthropology and Prehistory		
Email	patrick.sem@naturalsciences.be		
Telephone	+3226274380		
Website	www.naturalsciences.be		
Full name of Director General (DG) / President of the Institution	Patricia Supply		
Official function of DG / President	General Director <i>a.i.</i>		

Email of DG / President	direction@naturalsciences.be
-------------------------	------------------------------

## NON-ELIGIBLE PARTNERS CONTRIBUTING TO THE PROJECT CASH OR IN KIND

These are partners who are NOT paid by the project, and they do not appear in the previous section. Two tables are provided here, if there are more partners financed by the project, please replicate the table as many times as necessary, adding Partner C, Partner D...

### PARTNER A: [biodiversity.be]

International partner?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No	
Civility	<input type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input checked="" type="checkbox"/> Other: Ir.
First Name	André		
Last Name	Heughebaert		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Belgian Biodiversity Platform		
Department	Click or tap here to enter text.		
Service	IT & GBIF HoD		
Email	a.heughebaert@biodiversity.be		
Telephone	+32498088808		
Website	https://www.biodiversity.be/		
Full name of Director General (DG) / President of the Institution	Arnaud Vajda		
Official function of DG / President	President		
Email of DG / President	arnaud.vajda@belspo.be		

### PARTNER B: [MBG and DiSSCo Flanders]

International partner?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No	
Civility	<input type="checkbox"/> Prof.	<input checked="" type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Frederik		
Last Name	Leliaert		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other: [Specify]
Language	<input checked="" type="checkbox"/> NL	<input type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Meise Botanic Garden		
Department	Herbarium & Library		
Service	Click or tap here to enter text.		
Email	frederik.leliaert@plantentuinmeise.be		
Telephone	+32472301653		
Website	plantentuinmeise.be		
Full name of Director General (DG) / President of the Institution	Steven Dessein		
Official function of DG / President	CEO		
Email of DG / President	steven.dessein@plantentuinmeise.be		

### PARTNER C: [ULiège]

International partner?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No
------------------------	------------------------------	--

Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Frederic		
Last Name	Francis		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other: [Specify]
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Université de Liège		
Department	Functional & evolutionary Entomology		
Service	Click or tap here to enter text.		
Email	frederic.francis@uliege.be		
Telephone	+3281622283		
Website	<a href="http://www.gembloux.ulg.ac.be/entomologie-fonctionnelle-et-evolutive/">http://www.gembloux.ulg.ac.be/entomologie-fonctionnelle-et-evolutive/</a>		
Full name of Director General (DG) / President of the Institution			
Official function of DG / President			
Email of DG / President			

### PARTNER D: [UMons]

International partner?	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No	
Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Denis		
Last Name	Michez		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other: [Specify]
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input type="checkbox"/> EN
Institution	Université de Mons		
Department	Biology		
Service	Laboratoire de Zoologie		
Email	Denis.MICHEZ@umons.ac.be		
Telephone	+3265373807		
Website	<a href="http://zoologie.umons.ac.be/equipe_michez.htm">http://zoologie.umons.ac.be/equipe_michez.htm</a>		
Full name of Director General (DG) / President of the Institution			
Official function of DG / President			
Email of DG / President			

**PARTNER E: [CoEB]**

International partner?	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	
Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Beth		
Last Name	Kaplin		
Gender	<input checked="" type="checkbox"/> Female	<input type="checkbox"/> Male	<input type="checkbox"/> Other: [Specify]
Language	<input type="checkbox"/> NL	<input type="checkbox"/> FR	<input checked="" type="checkbox"/> EN
Institution	Center of Excellence in Biodiversity and Natural Resource Management		
Department			
Service			
Email			
Telephone			
Website	<a href="https://rbis.ur.ac.rw">https://rbis.ur.ac.rw</a>		
Full name of Director General (DG) / President of the Institution			
Official function of DG / President			
Email of DG / President			

**PARTNER F: [OBPE]**

International partner?	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	
Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Berchmans		
Last Name	Hatungimana		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input checked="" type="checkbox"/> EN
Institution	Ministry of Environment Agriculture and livestock		
Department	Burundian Office for the Protection of the Environment (OBPE)		
Service			
Email	hatungimanaberchmans@yahoo.fr		
Telephone	+257 69177962/79812477		
Website	<a href="https://bi.chm-cbd.net">https://bi.chm-cbd.net</a>		
Full name of Director General (DG) / President of the Institution			
Official function of DG / President			
Email of DG / President			

**PARTNER G: [CSB]**

International partner?	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	
Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Dudu		
Last Name	Akaibe Migumiru		
Gender	<input type="checkbox"/> Female	<input checked="" type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input checked="" type="checkbox"/> EN
Institution	Université de Kisangani		
Department	Centre de Surveillance de la Biodiversité		

Service	
Email	Duduakaibe@yahoo.fr
Telephone	+243823627533
Website	<a href="https://centresurveillancebiodiversite.org/">https://centresurveillancebiodiversite.org/</a>
Full name of Director General (DG) / President of the Institution	
Official function of DG / President	
Email of DG / President	

## PARTNER H: [University of Saint-Louis - Bruxelles]

International partner?	<input type="checkbox"/> Yes	<input type="checkbox"/> No	
Civility	<input checked="" type="checkbox"/> Prof.	<input type="checkbox"/> Dr.	<input type="checkbox"/> Other:
First Name	Marie-Sophie		
Last Name	de Clippele		
Gender	<input checked="" type="checkbox"/> Female	<input type="checkbox"/> Male	<input type="checkbox"/> Other:
Language	<input type="checkbox"/> NL	<input checked="" type="checkbox"/> FR	<input checked="" type="checkbox"/> EN
Institution	University of Saint-Louis - Bruxelles		
Department	Law Faculty		
Service	Research Centre for Environmental and Heritage Law (CEDRE)		
Email	<a href="mailto:marie-sophie.declippele@usaintlouis.be">marie-sophie.declippele@usaintlouis.be</a>		
Telephone	+32 2 211 78 24		
Website	<a href="https://www.usaintlouis.be/sl/4002929.html">https://www.usaintlouis.be/sl/4002929.html</a>		
Full name of Director General (DG) / President of the Institution	Pierre Jadoul		
Official function of DG / President	Rector		
Email of DG / President	<a href="mailto:pierre.jadoul@usaintlouis.be">pierre.jadoul@usaintlouis.be</a>		

## SUBCONTRACTORS

Two tables are provided here, if there are more subcontractors to be hired by the partners in the project, please replicate the table as many times as necessary, adding Subcontractor 3, Subcontractor 4...

### SUBCONTRACTOR 1: [INSERT ACRONYM OF THE INSTITUTION HERE]

Institution / Company	Belnet
Email	<a href="mailto:servicedesk@belnet.be">servicedesk@belnet.be</a>
Telephone	+3227903300
Website	<a href="https://www.belnet.be">https://www.belnet.be</a>

# PART C: DESCRIPTION OF THE PROPOSAL

## 2. COMPLIANCE WITH THE SCOPE OF THE CALL (MAX. 0,5 PAGES)

The project CANATHIST (The Natural History collections collected in central Africa by Belgian (federal) institutions) meets the INFRA-FED – impulse call objectives as it concerns the development of an open access platform that allows access to services like collection management system, IIF Images, 3D Files, maps , annotation, bibliographic servers and associated online/offline tools.

CANATHIST will therefore offer to the FSIs and other users/clients of Belnet a set of integrated open source tools/services allowing them to contribute to international infrastructures within the setup of EOSC and/or providing to overseas countries the FAIR access to the digital collections related to their own country including offline access and appropriate technologies according with the Access and Benefit-Sharing framework. Most of the tools exist already but are not proposed to the institutions/users in an integrated, sustainable and structural way. All this with the objective of stimulating the contribution of the federal collections and associated data to national and international research networks such as DiSSCo.

CANATHIST will contribute to the digital strategy of the federal government to implement an open data and open science policy to support the knowledge-based mission of the federal scientific institutions on the one hand, and to promote the preservation, management and valorization of the federal (analogue and digital) collections on the other.

The project will use as case study the collections coming from the colonial period which are possibly subject to a request of restitution. Therefore the project is very topical on the political agenda: on July 3 2022, the Federal Parliament adopted a [Bill Recognizing the Alienability of Goods Linked to the Belgian State's Colonial Past and Determining a Legal Framework for Their Restitution and Return](#) (hereafter 'Restitution Bill'). The collections in this project fall into the scope of the Restitution Bill. Restitution is also prominently present in the European and international debate, as more and more former colonizing countries are exploring avenues toward some form of restitution.

The developed platform could also be an important tool for the proposed Centre of excellence dedicated to the provenance studies allowing to centralize the data and metadata related to the collections and archives, and sharing the information with the different stakeholders representative of the countries of origin and the local communities.

## 3. SCIENTIFIC, TECHNICAL, TECHNOLOGICAL DESCRIPTION

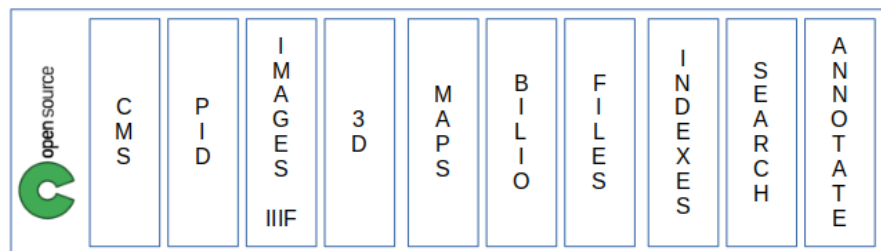
### 3.1. OBJECTIVES AND STATE OF THE ART



### 3.1.1 Project objectives and state of the art (max. 3 pages, references excluded)

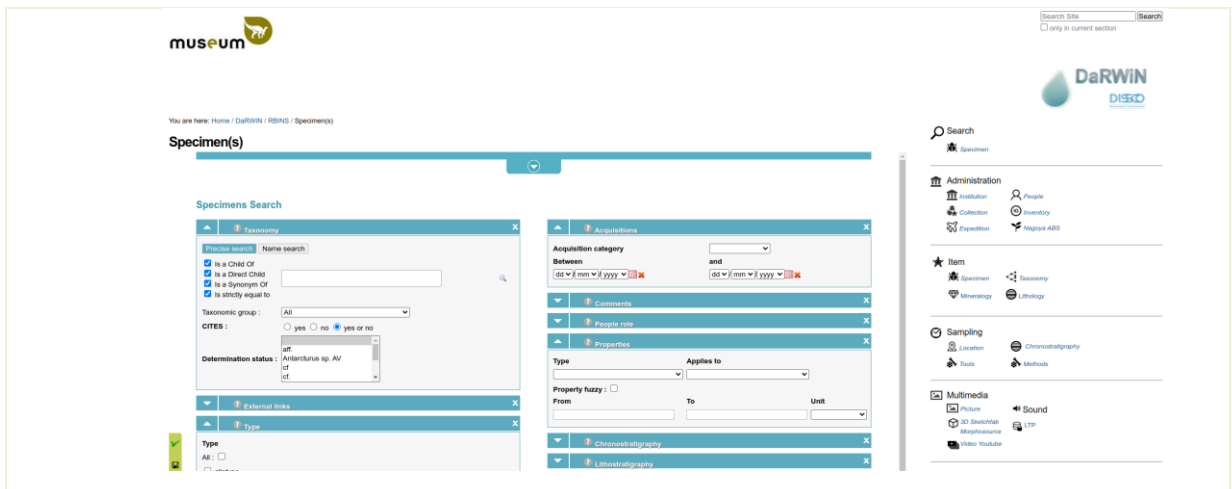
The project is articulated with two main objectives of equal importance:

- A) The main objective of the CANATHIST project is to build a portfolio of integrated open source tools/services for accessing the Belgian collections and evaluate how Belnet could help for the hosting of these tools for the FSI's and other Belnet clients. This includes Collections Management Systems (CMS), IIIF images server with dedicated viewer (e.g. Universal Viewer, Mirador), 3D files server with online viewer (e.g. 3DHOP), map server for distribution maps (e.g. geoserver & geonetwork), bibliography server (e.g. ORFEO), access to the BELSPO Long Term Preservation platform (Belspo LTP) and dedicated space/tool(s) for files exchange (e.g. Belnet FileSender) as well as global indexation system (e.g. ElasticSearch) and search interface (e.g. Symphony). Many of these tools have already been used/developed in the framework of previous projects but the integration of the tools should offer a better user experience, sustainability and access by the different FSIs and other Belnet clients.



The portfolio of Open Source blocks building the CANATHIST infrastructure

RBINS and RMCA have collaborated since a decade in the setup and the use of a common Collection Management System. The DaRWIn CMS is based on a PostgreSQL database and a Symphony/PHP user interface dedicated for Natural History collections. The database is able to manage millions of specimens and was recently reworked in a modular approach allowing to connect external tools/services and multimedia contents (BRAIN project naturalheritage.be, 2019-2022). Offline XLS templates with online access to validation tools were also developed allowing mass import of pre-existing data or recording data during fieldwork. Dedicated scripts were also coded to export the data to the GBIF repository using IPT. A IIIF image service was developed on the base of the Collective Access CMS allowing batch import of the image files and associated metadata. The DaRWIn user interface was also adapted allowing the integration of the DaRWIn views in existing institution websites using Iframes.



Integration of the DarWIN views in the NaturalHeritage Portal (Plone CMS)  
The right menu is designed directly in Plone with a WYSIWYG editor

Unfortunately, the IT capacities / infrastructures of the RMCA and RBINS are limited and it is difficult to propose the DarWIN ecosystem as a sustainable service to the other Belgian institutions in the existing data centers. Moreover, the long term funding opportunities at the federal and regional levels are not optimized to promote a multi-level collaboration. This project will evaluate with the help of Belnet how to build a sustainable hosting of the “ecosystem” with appropriate business plans. Belnet already offers a set of tools and cloud services to the Belgian scientific community as well as the public administration. It is therefore the adequate support to deploy an integrated portfolio of (open source) tools related to the creation, conservation and valorization of the digital heritage linked to the collections.

The developed infrastructure could help other FSIs and/or other Belnet clients to contribute to the international network as ESFRIs (e.g. DiSSCo) and align the FSIs capacities with the development of the European Science Cloud (EOSC). It could also be the information infrastructure for the setup of the **Centre of Excellence for provenance research** recommended by [Restitution Belgium](#) and the BRAIN 2.0 [HOME project](#).

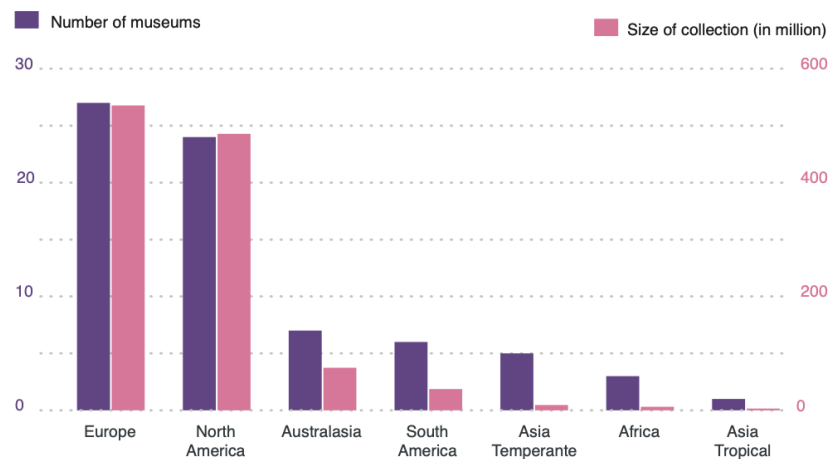
B) Case study: the Natural History collections from Central Africa housed by Belgian institutions

The case studies of the CANATHIST project are the Natural History collections mainly collected by the FSIs in Central Africa, during the colonial period. These samples/specimens were often divided between the Royal Museum of Central Africa (RMCA), the Royal Belgian Institute of Natural Sciences (RBINS) and the Meise botanic garden (MBG). After the independence of the Congo, a part of the collection was also prepared for repatriation and stored in the RMCA, still waiting for a formal agreement with a host institution in DRC. In 1961 it was decided to prepare a fourth of the entomological collection collected in the national parks of Congo for a restitution to RDC. About 20 years of work was necessary to sort out the specimens and prepare more than 6000 boxes. Since 1980, the material has been ready for restitution to African institutions, but has never been completed. Forty years later the boxes are still at the Africamuseum, but no longer have been curated by the museum(s).



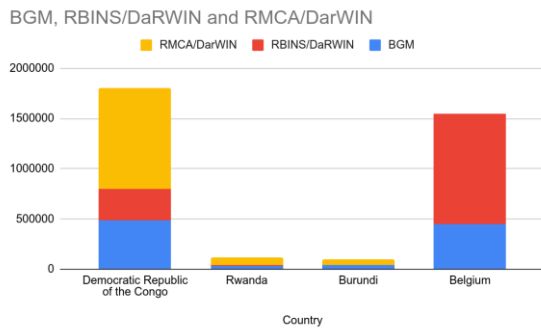
Collections prepared for restitution at AfricaMuseum

Africa houses a large part of the terrestrial biodiversity but suffers from the lack of African Natural History collections and is under-represented at the global scale .

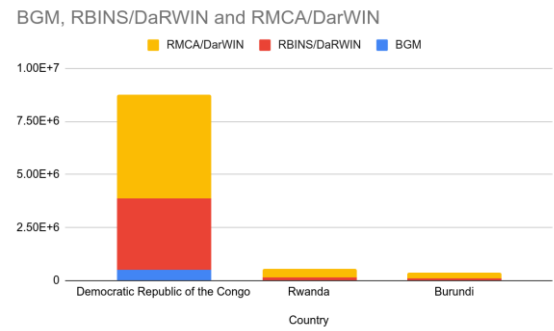


Ranked histogram of number of museums and their collection sizes, grouped by geographic region.  
From Johnson *et al.*, submitted to Science, Supplementary Fig. 1.

The Belgian federal collections (housed by RMCA, RBINS and MBG) are together one of the most important collections related to Central Africa worldwide. More than 1.5 million of specimens related to Central Africa are already encoded in the Institutions collections management systems. The evaluation of the total number of specimens collected in DRC, Rwanda and Burundi in the FSIs is evaluated to 9 millions on the basis of the proportion of the total number of specimens (48 millions) and the global ratio of digitization.



Number of specimens encoded in the CMS of RMCA, RBINS and MBG



Evaluation of the number of specimens housed by RMCA, RBINS and MBG

The Belgian federal collection related to Central Africa is therefore a unique heritage which has to be unlocked for the scientific community and the countries of origin (DRC, Rwanda and Burundi). The FAIR access to the data and metadata associated with these collections is required to allow these countries to manage the biodiversity inventories and the development of sustainable exploitation pathways of the natural resources.

The access by overseas users also has other user requirements than the “full online access” promoted by Belgian users, and hybrid scenarios have also to be developed in order to optimize the access / edition of the data by the DRC, Rwanda and Burundi stakeholders.

### 3.2. TRANSLATION OF THE OBJECTIVES INTO APPROPRIATE AND WELL-DESCRIBED METHODOLOGY (MAX. 10 PAGES, FOR 3.2.1 AND 3.2.2. SUB-CRITERIA)

#### 3.2.1. Methodological approach

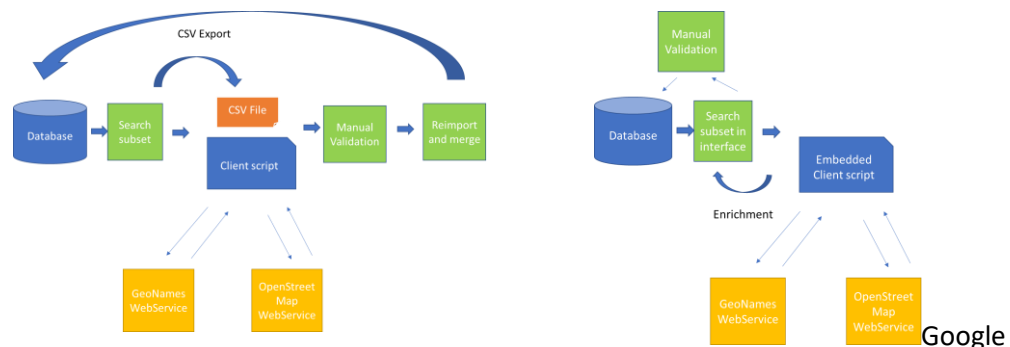
The idea of this project is not to develop new tools *de novo* but if possible to use/connect existing, but still state of the art, open source solutions for their applicability for managing federal and non-federal collections (specimens, objects, archives, multimedia).

In the past five years, RMCA and RBINS have been involved in the development and usage of several softwares for the management of their biological collections. One of the main goal of the NaturalHeritage project was the improvement of the integration of Darwin in scientific workflows, by developing modules to extract and distribute its own data (reporting in CSV and excel PDF formats, labeling, statistical summary and collection, generation of documents for loans and Nagoya) and import external data with a quality check mechanism, working in a semi-automatic way, via a

template (simple data model in spreadsheet format mechanism). Lightweight tools, compliant with this template structure have been developed in Excel and Open-Office to manage offline smaller collections (targeting African institutions) and prepare templates for import in the main DarWIN database and link them to image servers (in IIIF format) and external geographical databases (OpenStreetMap and GeoNames) to enrich them. Still, these tools currently lack a global integration, and the methodology to use them could be better formalized and documented, especially in the prospect of data restitution and FAIR ecosystems.

The main technical aspects of CANATHIST would focus on 3 objectives:

- improve and formalize the methodology to better trace the history of data (especially regarding their condition of acquisition), assess the sensitiveness of data, its complementarity with the scientific and curatorial needs of our central african partners, by lining these systems to external and cross-disciplinary sources
- merge together on a single redistributable platform (e.g. Docker & Docker Desktop) the existing packages that are currently on different platforms, and lower their technical complexity so that they could be installed and maintained in a small data-center installed in Central Africa or by a Belgian Institution with limited IT resources, or even used on a single computer with limited and/or non-permanent internet connection. This would imply automating the configuration of these tools.
- develop integrated systems for the enrichment of data (e.g. linking geo coordinates to current and (pre)colonial names). The enrichment and georeferencing is very important for provenance studies and a simplified and robust workflow has to be developed without having to tediously export batches of data and (re)import them after (re)validation):



current workflow

CANATHIST simplified workflow

As case study, the project will produce the most complete inventory of the Central Africa Natural History specimens and create a research infrastructure available for the DiSSCo community and the Central Africa stakeholders and linking with existing resources and data authorities:

- Preparing the provenance study, A single place with unique geographic coordinates can be named by several names that

change over time and languages, as is the case in the naming of places in DRC and Rwanda before, during and after the Belgian colonial period. The correct understanding of place names is therefore very important for provenance studies of cultural and natural collections. CANATHIST will investigate possible solutions for implementing multilingual and versioning mechanisms in online platforms. Correspondence between names of localities and geographic location will provide mapping data for distribution maps required for the provenance evaluation.

- Digitizing the collections (data, metadata and digital specimens). The project will inventorize all specimens and related data not yet computerized. This will include all basic primary information regarding the object's (specimens) identity (i.e. the information as provided on the associated labels) and their conservation status (e.g. intact, damaged) as well as high resolution digitization of important specimens (e.g. type specimens). Prioritization of the digitization will be established with African partners and according to the GBIF priorities and standards, new data/metadata will be integrated in the DaRWIN ecosystem used in both institutions
- Sharing all information, digital specimens and archives with the competent authorities, institutions and citizen communities of DRC, Burundi and Rwanda thanks to a specific collaborative platform. This also includes connectors to crowdsourcing platforms such as the [DOEDAT](#) platform hosted by the MBG or the international [Zooniverse](#) platform.
- Developing visualization and 3D measures tools allowing scientists to produce advanced scientific analysis on digitized specimens. During the last decades, technological developments in digitization allowed the production of 3D virtual specimens for scientific and education purposes. The approach is mainly based on the creation of a 3D model representing the specimen and allowing complex 3D measurements. Nevertheless, this approach has limitations if the scanned surface is composed of thin and reflecting material largely represented in natural history collections (e.g. insects wings, feathers, hairs, leaves). The CANATHIST proposal will develop a specific viewer allowing to take 3D measures on the base of a set of oriented 2D images. This approach also has the advantage that it does not require the computation of the 3D model which needs a lot of computing time.

The digitization itself is an important part of the CANATHIST infrastructure but is not financed by the proposal. The case study will benefit of the synergy with other structural funding of RMCA and RBINS:

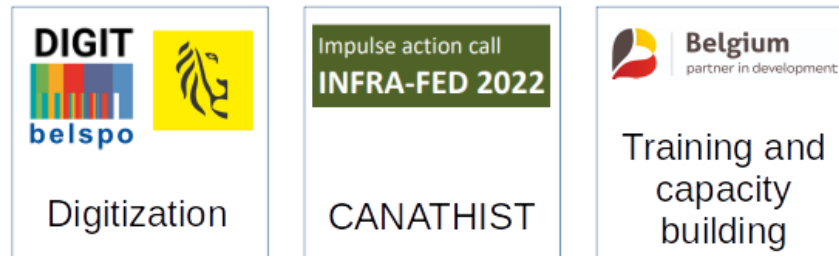
- Prioritization of the [DIGIT-4 program](#) will be given to the specimens collected in Central Africa allowing the digitization of the specimens and associated data and metadata
- Reorganization of the wet collections at the Africa Museum with associated digitization
- Massive disinfection/cleaning of the bird collection at the Africa Museum with associated digitization

The case study will also reuse digital collections from previous digitization projects:

- about 2 million of specimens already encoded in the RMCA and RBINS DaRWIN systems

- More than [500.000 herbarium sheets](#) digitized by MBG (DOE and DOE2)
- About [70.000 pictures](#) were produced during the fieldworks of the Parcs nationaux du Congo and already digitized by specific projects at RBINS.

Specific training and capacity building will also be organized by RMCA and RBINS with the partners of RDC, Rwanda and Burundi using the development cooperation framework agreement (DGD).



The 3 funding pillars of the proposed infrastructure

During previous projects funded by the DGD (BICS, CABIN) we noticed that the technical and scientific user requirements are very similar between Belgium and Central Africa. However, the profiles and resources of the scientists greatly differ:

- Central African scientists are often younger, at the beginning of their career and/or engaged in Doctoral studies. Besides the lack of institution resources, they also have to spend a large amount of time on field trips and redacting their thesis as well as scientific articles without fast internet connection. Even if they have a relevant approach of the needs and technical aspects of the management of natural history collection, they are rarely in position to implement this at institutional level.
- Several institutions (e.g. the ichthyological faculties of the Université de Lubumbashi and the Institut Supérieur Pédagogique de Bukavu, in the field of ichthyology) gathered sizable physical collections and build teams with identified technical staff. They are also asking for strengthening their technical skills in terms of database management.

CANATHIST will therefore contribute to produce, document and link a suite of softwares, that will be packaged together and deployed both in Belgian ESF (and/or in a cloud) and in smaller data centers in Africa, using Open Source technologies.

Besides of this, the model of the current CMS (e.g Darwin for biological collections and Collective Access for images and representation of cultural objects) should be enlarged and linked to web services to handle both the provenance of objects (in the prospects of collection restitution) and the legal framework of the current collections (e.g Nagoya).

CANATHIST will also do a model analysis, involving the central African partners, aiming at a technical convergence and standardization to publish and exchange information at international level (e.g. GBIF, Darwin Core) and improving the FAIR access of the data and metadata.

The user-documentation produced by the project will be 3-folded targeting:



- IT administrator having to install the software,
- Collection technician using it,
- Scientists managing them

It will also encompass the workflow adding extra information on the provenance and legal status (in terms of ownership and/or guarantee of public access) of both the collection objects and (meta)data describing them.

### 3.2.2 Methodology

Please consult the  *Gender checklist* when filling out this part of the proposal and fill out the  *Ethics form*.

The CANATHIST methodology is translated into five main work packages (WP).

The **WP1** and **WP2** are the project and data management work packages. They will ensure that the project will produce the required outputs and that the data produced by the project will be saved and available according to the Data Management Plan.

The **WP3** will prepare a survey to make the most complete inventory of Natural History Collections from central Africa housed by the Belgian and Africa Partners. This inventory does not require a complete digitisation of the collections but will allow the partners to prioritize the digitization efforts of the collections housed in Belgium following the requirements of the African Partners as well as the existing priorities established by the international community (GBIF, DiSSCo). We will use the Collection Registry survey developed by the CETAF and the DiSSCo Prepare WP 8.1 Specialization Plan as base of the survey. The results of the CANATHIST survey will be therefore compatible with these exercises but with a higher level of precision. The WP3 will also contribute to the provenance study establishing the geographic origin of the collect / sampling location establishing a multilingual catalog of the names of localities in Central Africa taking account of the modifications of the names before, during and after the colonial period.

The **WP4** will focus on the ethical and legal aspects. For the specific matter of natural history collections, several legal frameworks are applicable. First, natural history collections entail genetic resources and therefore fall under the scope of the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits, which entered into force on 12 October 2014 and is part of the CBD. This international agreement aims at sharing the benefits arising from the utilization of genetic resources in a fair and equitable way (<https://www.cbd.int/abs/>). The Nagoya Protocol is focussing mainly on genetic resources (which can also be seen as a kind of digitization) to protect the countries and communities of origin. It has also a significant impact on the non-commercial scientific institutions which need specific agreement to collect and study new specimens even for non commercial purposes. The principle of Fair and Equitable Sharing of Benefits is defined by a set of monetary and non-monetary benefits. Some of them can be linked to the digitisation strategy and the (co-)ownership of the resulting digital data:

- Joint ownership of relevant intellectual property rights;
- Collaboration, cooperation and contribution in education and training;
- Admittance to ex situ facilities of genetic resources and to databases;
- Strengthening capacities for technology transfer;



- Access to scientific information relevant to conservation and sustainable use of biological diversity, including biological inventories and taxonomic studies;
- Institutional and professional relationships that can arise from an access and benefit-sharing agreement and subsequent collaborative activities

Second, natural history collections also fall under the scope of heritage legislation. By contrast with genetic resources, very few international legislation exists concerning the access to these collections, and specifically their digital access. Indeed, The UNESCO “[Charter on the Preservation of Digital Heritage](#)” (2003) and the “[Recommendation of UNESCO Concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form](#)” (2015) are the only two soft law instruments addressing the issue, but they ignore most other set of norms interacting with these issues (intellectual property rights, digital rights, data protection, information technology...). In other words, there is a legal void for addressing digital and digitized cultural heritage. In Europe, some regulation exists however, mainly under the auspices of the European Union. The European Commission even adopted a [Recommendation on a common European data space for cultural heritage](#) (10 Nov 2021) and granted a [tender on this project to Europeana](#), the EU’s digital platform for heritage. These international and European norms are very relevant for the proposed project, even if these norms are still partially insufficient (some regulations are still in process of adoption or don’t concern the present topic) or if some are not binding (recommendations).

Restitution of such digitized cultural heritage is even more a blind spot in the law. It has not really been addressed in the Sarr-Savoy Report, or at least not in the right nor sufficient way like Pavis and Wallace commented in their response (2019). The German Report on Colonial Collections makes a clearer attempt and will be closely looked into. Yet, these attempts intertwine legal and ethical considerations - given the legal void - in order to offer a satisfying response to the issue. This state of the arts highlights the need to have a thorough research on the legal aspects of the digital side in the restitution debate.

In the report on the state of the art for legal frameworks concerning natural history collections, some ethical and more “meta” questions will also be raised as both aspects interact.

The first important set of issues concern the **ownership** or absence of such ownership on the digital derivatives of natural history collections and more broadly the legal status given to these digital derivatives. In other words, who owns the data generated by the digitization of natural history collections? Is it even to be owned as most scholars contest data being subject to property right, considering it rather as a common good<sup>1</sup>? And what about intellectual property rights on these digital derivatives?

Data in itself is indeed generally accepted as extrapatrimonial, out of commerce so to say. Yet the commodification of data, and of heritage data, is an economical and social reality<sup>2</sup>. It could pave the

---

<sup>1</sup> A. Strowel, « Omnia sunt ©ommunia : des opera au Big Data », *Revue interdisciplinaire d’études juridiques*, 2018/2, pp. 177-209.

<sup>2</sup> N. Silberman (2014). From Cultural Property to Cultural Data: The Multiple Dimensions of “Ownership” in a Global Digital Age. *International Journal of Cultural Property*, 21(3), 365-374. doi:10.1017/S0940739114000162

way towards a “meta-cultural property that represents a shared global culture that we are creating today”<sup>3</sup>.

Moreover, a set of data can benefit from specific intellectual property rights and copyright may also apply in certain cases. However, one may ascertain that in most cases, unless these 2D or 3D or otherwise digitized heritage are original in themselves, they are not protected under copyright law, nor subject to any other intellectual property rights, except for the software used (if sufficiently new and inventive there may be a patent, or if the coding follows an original writing in an artificial language, there may be copyright).

The data of these digital derivatives are subject to a certain form of data protection legislation. As such, the [Digital Single Market Directive](#) (transposed by the Belgian [law of 19 June 2022](#)) offers the best framework to date to deal with digital heritage. Moreover, the [Directive \(EU\) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information](#) was transposed in the [Belgian Law of 14 May 2016 on the re-use of public sector information](#) (modified in 2021) and contains relevant definition (of research data, dynamic data and re-use) that participate in the existing legal framework, but not much more. Also, the [Regulation 2022/868 on European data governance \(Data Governance Act\) of 30 May 2022](#) gives a broad definition of data as “any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording;” (art. 2.1.). Finally, some specific regulations for (cultural) platforms displaying these data will be analyzed, such as the [Digital Services Act](#) and the [Digital Market Act](#), raising the issue of open access.

This proposal aims to **conduct a thorough analysis of the existing regulations on the matter, mostly at European and national level, together with scholar interpretations surrounding (data) ownership.**

Closely linked to the notion of data ownership is that of **data sovereignty**<sup>4</sup>, showing the difficulty to think in terms of appropriation but still reflect the idea of exclusivity and control.

This leads to a second aspect besides the ownership issue: who gets to decide about the digitization of natural history collections ? Who **controls** the process of digitization? Indeed, as no ownership rights might apply, user’s rights are nonetheless applicable (users such as cultural institutions, digitizing companies, owners of the material support, countries, institutions and communities of origin ...) and careful **governance models** should be drafted in order to determine which party may claim which rights, in the context of commodification and even financialization of data. Specifically with regards to the restitution debate, and particularly in the postcolonial context, the issue to handle digitization in culture-specific ways is essential. Therefore, ethical considerations must cross legal analysis when determining what legal framework (Belgian, European or African? According to the place of digitization or the place of restitution?) should be applied; when exporting legal notions such as public domain and open access to foreign legal orders; or when determining the value of the digital derivatives (social and economical)<sup>5</sup>.

---

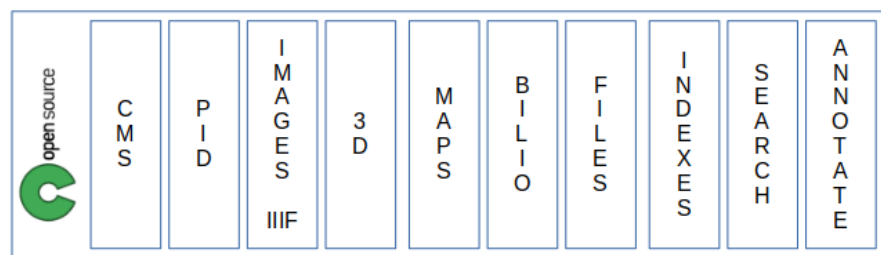
<sup>3</sup> N. Silberman (2014). From Cultural Property to Cultural Data: The Multiple Dimensions of “Ownership” in a Global Digital Age. *International Journal of Cultural Property*, 21(3), 365-374. doi:10.1017/S0940739114000162

<sup>4</sup> T. KUKUTAI et J. TAYLOR (dirs.), *Indigenous Data Sovereignty*, 38, Australia, ANU Press, 2016

<sup>5</sup> L. Lixinski, « Digital Heritage Surrogates, Decolonization, and International Law: Restitution, Control, and the Creation of Value as Reparations and Emancipation », *Santander Art and Culture Law Review (SACLAR)*, 2020, vol. 2020, n° 2, pp. 65-86.

Third, the notion of **access** itself should be examined, including from an ethical point of view: the desire to grant universal access to certain dematerialized natural history content may come up against certain rights and interests, particularly those of the communities of origin, let alone when it comes to digitizing sacred objects. Following a decolonial approach to the concepts of access and dematerialization, thought should be given to the inclusion of these communities in the digitization process as well as in the access policies of these digital contents.

The **WP 5** will combine the tools and services using several Open Source blocks which together build the CANATHIST infrastructure.



The building blocks of the CANATHIST infrastructure

### **CMS: The Collection Management System**

Online Collection Management system(s): DaRWIn or Collective Access for cultural Heritage

At the level of the two FSIs partners, the main collection management system is DaRWIn (Data Research Warehouse Information Network). DaRWIn is an open source system, consisting of a [PostgreSQL](#) database and a customizable web-interface based on the [Symfony framework](#). DaRWIn is an in-house solution developed by the Royal Belgian Institute of Natural Sciences (RBINS), as a Natural History collections management system for biological and geological samples. In 2014, the Royal Museum for Central Africa (RMCA) adopted this system for its collections and started to take part in new developments. DaRWIn stores sample data and related information such as place and date of collection, missions and collectors, identifiers, technicians involved, taxonomy, identification information (type, stage, state, etc.), bibliography, related files, storage, etc. Other features that deal with day-to-day curation operations are available: loans, printing of labels for storage, statistics and reporting. The DaRWIn database currently manages information on more than 600,000 records (about 4 million specimens / 38 million specimens) housed at the RBINS and more than 650,000 records (more than 1 million specimens / 10 million specimens) at the RMCA. DaRWIn is accessible online. A [Github repository](#) is available for reuse by the community.

DaRWIn is an online CMS but several offline tools were developed to help collection managers and scientists to prepare/encode the information offline using Open Source applications:

- Offline collection management tools using spreadsheets templates (XLS or ODS)

Specific templates were developed in the framework of the NaturalHeritage BRAIN project. XLS and ODS templates were prepared to fill data before mass import in DaRWIn. Specific validation tools were developed to check taxonomy and geographic names. These templates can be used on laptops (Windows and Linux) with macro using (visual)basic language. This

approach allows the development of custom templates on demand following the user requirements of scientists for offline encoding during the fieldwork campaign.

- Offline / online data collection (Kobo collect & [Kobo Toolbox](#))

A second option which is already used at the Africa Museum is the Kobocollect android app. It allows you to encode forms using the tools of the android devices (camera, gps, barcode reader). Kobocollect was initially developed for the data collection of humanitarian organizations, but is available for other research areas. The system is working directly online or offline with ssl synchronization when internet is available. It is possible to use the main server for small projects (<10,000 Submissions/month) or use a local server using docker for larger projects.

### **PID server**

The unique identification of digital information is of first importance especially if the infrastructure is running online / offline allowing to identify and connect the data linked to a specimen.

DaRWIN is already using stable identifiers as promoted by [CETAF](#). Existing identifiers as [ROR](#) for Institutions, [ORCID](#) for scientists, [GRSciColl](#) for collections, [DOI](#) for publications and datasets, Geonames ID already facilitate the unicity of information but it is interesting to manage a PID server. Belnet is looking for interest in Belnet's development of a 'PID' (Persistent Identifier) system. A PID is a unique digital identifier given to a dataset so that it can be identified at all times, even with changes in location on the Internet. Thus, assigning a PID contributes greatly to the findability of data – one of the FAIR principles.

### **Images server 2D & 3D**

The value of Natural History collections does rely on the specimen itself but also on all the associated documentation. An important part of this information is linked to multimedia data (pictures, 3d files, sounds and videos) or to the associated literature and archives. Access to the digital representation(s) of specimens/objects is a fundamental requirement for many research and citizen activities. CANATHIST will investigate and set up specific servers/viewers allowing a rich, online viewing environment for digital objects to be presented to the platform users.

- **2D Images server** ([Collective Access](#), CA)

Collective Access is a free, open-source software for cataloging and publishing museum and archival collections. It is already used by several Belgian Institutions:

- KULeuven uses CA for the management of the collections.
- RMCA is using CA to provide access to the 2D and 3D virtual collection(s) using faceted search and IIF image server.
- RBINS uses CA CMS as an image server.

One of the main advantages of CA is the possibility to import images and metadata (of images and/or items) using XLS templates. The images can also be integrated in other Collection Management Systems like DaRWIN for the natural history collections using IIF viewers like Universal Viewer or Mirador.

- **3D objects server** (3DHOP) and RTI server (Relighth)

Most Belgian institutions are presently using the commercial Sketchfab platform which is certainly the cheapest solution to share high resolution 3d models on the internet. DaRWIn has a specific view for 3D objects stored on sketchfab. Nevertheless, the licensing and user conditions can change with time and a more structural option has to be developed for the future.

[3DHOP](#) is an open-source framework for the creation of interactive Web presentations of high-resolution 3D models, oriented to the Cultural Heritage field. Relight is a library to create and view (on the web) relightable images (RTI). Both are used in the Visual Media Service, developed in the framework of several EU projects including the H2020 EOSC pilot. It was developed by [ISTI, CNR](#) which is also an important partner of the Robotics for Digitisation (r4d) proposal submitted in november 2022 which aims to set up the robotic digitization of European collections. The Smithsonian X 3D Explorer will also be evaluated in the framework of the DIGIT-4 program.

### **Maps server and distribution maps**

The integration of georeferencing tools, linking source data of the partner institutions to web APIs for geographical data (e.g. OpenStreetMap, Geonames, NaturalEarth, Marine regions, NIMA, Wikidata and Getty Thesaurus), will enrich the partner institutions' data with coordinates, and supports multilingual extension of place names with historical versioning.

The integration of these georeferencing tools will be develop in a four step approach:

1. A central GIS node, likely based on PostGIS, will harvest the contributing databases (Darwin, ...), or be available for push if the contributing sources have no API.
2. The GIS database contains an embedded script to query the reference databases. Data (metadata and coordinates) are sent back to the GIS database and stored in a table linked to the original values.
3. A data manager or database administrator performs the manual validation of the data via an online graphical interface. Validated data is sent back to the original database.
4. Validated data is exposed to two publicly available data:
  - georeferenced data will be published on a WMS/WFS server and to Web APIs (Geoserver)
  - metadata will be published on a CSW catalog (GeoNetwork) which is compliant with the European [INSPIRE](#) directive. It also allows users to annotate the metadata with keywords (including the multilingual [GEMET](#) repository).

### **Bibliography & Archives**

The DaRWIn CMS can link bibliography to specimens using internal bibliography or hyperlinks. The CANATHIST will improve the integration with 2 Bibliography repositories using OAI-PMH protocol:

- [ORFEO](#)  
**ORFEO** is the institutional Open Access repository for Federal Science Policy funded research. It is based on [DSpace](#), an Open Access product building Open Digital Repositories. Orfeo is

hosted by Belnet. Orfeo uses [OAI-PMH protocol](#) to gather the information. The CANATHIST infrastructure will use the OAI-PMH connector to exchange information with the Orfeo repository. The OAI-PMH will also be used to connect with other repositories such as SODHA, the Belgian federal data archive for social sciences and the digital humanities.

- [biblio.naturalsciences.be](#)

This bib4plone addon of the Plone CMS is used by RBINS to store and share the bibliography, the PDFs (Open Access) produced by scientists of the RBINS and/or edited by RBINS as (co)editor. This system today includes about 30.000 bibliographic references compatible with BibTex and RIS format and 20.000 associated PDF files including the publications related to the field works of the [Institute of the National Parks of Belgian Congo](#)".

- The proposal will also evaluate how to use/integrate the open source [Zotero](#) bibliography manager in the CANATHIST infrastructure.

### **Files exchange**

An important part of the CANATHIST infrastructure is the capacity to send / exchange large files which are not compatible with emails or free google drive. This will be certainly the case of the packed infrastructure available for transfer to Central Africa partners. The size of the large files is also a limitation for Belgian Partners as some digital datasets can easily be larger than 200 Gb (e.g.  $\mu$ CT datasets). The CANATHIST infrastructure will use the [FileSender](#) server proposed by BELSPO.

FileSender is available via a simple interface which enables sending very large files, up to 5 TB. FileSender is available to the member organizations of Belnet's R&E Federation at no additional charge. Filesender will be used for the transfer of dockers with specific blocks and/or datasets.

One development of the FileSender service could be the direct link between the Belspo Long Term Preservation Platform (LTP) and the Belnet File sender service as an output of the LTP retrieve process.

### **Indexes and Searches**

The indexes are the main entry points into the CANATHIST infrastructure, as they can align data coming from different sources into a common model and web API. Several technologies are already in use in the FSIs for this purpose, especially the Elasticsearch technology which was developed in the framework of the [Naturalheritage](#) project. Elasticsearch indexes can produce fast answers to faceted search of very large amounts of data. The creation of Elasticsearch indexes allows to aggregate several datasets from different (re)sources and create cross-disciplinary searches.

The search interface will (re)use the NaturalHeritage results. User interfaces with faceted criteria will be developed as simple views using a web development framework (e.g. Symfony, Django) to enforce a layered architecture and ease intervention of tier parties and maintainers. The views can be integrated in institution/project websites using Iframes.

The CANATHIST proposal will also evaluate graphical user interfaces such as the [DEJAVU Web UI](#) for Elasticsearch. DEJAVU is available with an open-source MIT license and allows users to browse and import data of Elasticsearch indexes, prepare searches and display the results of the searches. This product can be used as a Chrome extension or inside of a Docker. The graphical interface facilitates the creation of ad hoc search views and reporting tasks.

The task will also enforce a security layer that is not featured in the free versions of Elasticsearch (e.g. by evaluating [OpenSearch](#), a fork of Elasticsearch that natively embeds SSL protection and user accounts with roles and access levels).












## Annotation

The annotation features are very important in the setup of such infrastructure.


















Several tools provide annotation capabilities as the [Mirador IIIF viewer](#) but the system has to be linked to an annotation server to save the annotation and it is not possible to share them with the community without a specific annotation server. No annotation server was installed during previous projects. The CANATHIST will test the Mirador IIIF viewer with an annotation server in the proposal. We will also evaluate [Label Studio](#) as a general annotation tool. This Open Source Data Labeling Platform allows to annotate images, audio, text, time series and video. It can also be installed with Docker and can be connected using API. It is therefore an interesting candidate for the integration in the CANATHIST infrastructure.

The developed infrastructure has also to be connected to external crowdsourcing as the [DOEDAT](#) platform hosted by the MGB, the [Zooniverse](#) or to any other crowdsourcing platform installed by a FSI (e.g. [PyBossa](#))

The following table summarizes the technical architecture of the CANATHIST infrastructure using Open Source product and integration using Docker installations.

CANATHIST blocks	Main block	Alternate and/or additional
<b>Collections Management System</b>  Export Darwin Core		
<b>Offline editor</b> <i>XLS Templates and Android app</i>		
<b>PID</b> <i>Belnet service</i>		
<b>Images server</b> 		
<b>Images IIIF Viewer</b> 		
<b>Images and 3D measures</b>	<b>CANATHIST</b>	
<b>3D</b> 	<b>3DHOP</b>	




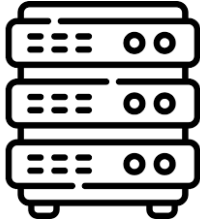

<b>Maps</b> 		
<b>Bibliography</b>		
<b>Open Archives</b> OAI-PMH		
<b>Larges Files</b> <i>Belnet service</i>		
<b>Indexes</b> and export API 		 SSL
<b>Searches Web UI</b> 		
<b>Annotation</b> 		
<b>Crowdsourcing</b> <b>Connectors</b>		

The proposed building blocks of the CANTHIST infrastructure

The Docker installation of the different tools allows the CANATHIST infrastructure to operate with different conditions of internet access and different hardware configurations.

<b>Full time High speed internet connection</b>	<b>Part-time High speed internet connection</b>	<b>Part-time Low speed internet connection or no connection</b>
---	---	---



		
<i>Cloud servers &amp; storage</i>	<i>Local servers &amp; storage</i>	<i>Standalone workstation</i>

The possible hardware configuration of the CANATHIST infrastructure

This approach opens the use of the CANATHIST infrastructure in many different internet access conditions including partners and users in the Central Africa countries where the internet could be limited.

The **WP6** of the CANATHIST proposal is related to the training and capacity building associated with the Infrastructure and the case study.

The proposal will prepare an extensive documentation of the CANATHIST infrastructure and associated blocks. This will help the different categories of users (e.g. IT, technicians and scientists) to use the infrastructure as active users or just to make searches and exports.

The [CEBioS](#) – ‘Capacities for Biodiversity and Sustainable Development’ housed at the Royal Belgian Institute of Natural Sciences (RBINS, BioPoIS group Belgian Biodiversity Policy Support Group) is funded by the Directorate-General for Development Cooperation (DGD). It will collaborate with the African partners and prepare specific training and capacity building activities. The Africa Museum will also play a role in this WP with the CoEB through several projects also funded by the DGD.

The **WP7** will explore with partners and Belnet how to prepare the future of CANATHIST itself but also how the infrastructure can be (re)used and extended by other FSIs or scientific institutions. The topic of the proposal and the case study is linked with the core activities of the Africa Museum and the RBINS concerning the natural history collections and the setup of the DiSSCo ESFRI. Nevertheless, it is not the role and in the IT capacities of the 2 FSIs to host the infrastructure for other institutions or topics. The integration of the building blocks of the infrastructure will stress on the capacity to host the infrastructure on institution servers, on the cloud or even on a single workstation. The 2 FSIs will analyze with Belnet how to offer this portfolio of tools and services to the other scientific institutions in a sustainable way using services hosted by Belnet or using the Belnet OCRE Cloud Catalogue. This will allow defining several scenarios with associated costs to the CANATHIST partners and other FSIs or other scientific institutions.

The **WP8** is dedicated to the dissemination activities. The CANATHIST infrastructure will be presented in (inter)national workgroups/associations/infrastructures as CETAF, DiSSCO, DIGIT. All developed tools and services will be available for (re)use on a specific branch of the naturalheritage Github.

## 4. IMPLEMENTATION & BUDGET

### 4.1 METHODOLOGICAL COLLABORATIVE APPROACH (MAX. 10 PAGES)

#### 4.1.1 Expertise in the frame of the project (max. 2 pages)

The different partners associated in the proposal have a strong experience in the proposed development :

- The **Africa Museum** is one of the World leading scientific institutions for the Central Africa area. The institution has also a huge experience in the development of technical / IT tools oriented to the user requirements in collection management, GIS, digitization. The institution also developed several projects related to the proposal in collaboration with the overseas countries and especially with DRC, Rwanda and Burundi.
- **RBINS** is one of the largest collections of natural history in Europe. RBINS developed the DaRWIn CMS for more than 15 years and coordinated/participated in several projects related to the digitization of the collections. RBINS houses the Collections and scientific archives of the “Parcs nationaux du Congo”. The [CEBioS](#) has a strong position working with people, for people and nature, by facilitating best practices, skills, tools and knowledge in the field of biodiversity and sustainable development. At the international level, CEBioS operates in the framework of the obligations of Belgium and the partner countries to the Rio Convention on Biological Diversity (CBD) and related protocols.
- The **Meise Botanic Garden** is the third institution managing the Belgian federal collections collected in Central Africa during the colonial period. MBG participates in several international networks and infrastructures and developed several scientific projects and trainings in Central Africa. MBG is also the coordinator of DiSSCO Flanders.
- The **University Saint Louis** is the partner of the HOME project related to the provenance study of the Human remains hosted by FSI and developed a strong expertise in legal and ethical aspects related to the questions of restitution of cultural/natural heritage objects covered by the Law of 3 July 2022. Marie-Sophie de Clippele has also interest in the legal and ethical questions related to the digitization of objects collected without the consent of local communities and to the sharing of these digital collections. She is co-editing a special issue on these questions for the International Journal of Semiotics of Law.
- The **Biodiversity Platform** has a large experience in the data validation and publishing of data related to the natural sciences and to biodiversity. The Belgian Biodiversity Platform provides services to the Belgian scientific community engaged in biodiversity research and to policy-makers as well as to practitioners. It provides services related to the mobilization, publication, access and use of biodiversity data in an open and free manner.
- The **ULiège** and the insectarium HEXAPODA is the third largest entomological collection in Belgium. Several of the collections housed by the institution are related to the Central Africa and the **ULiège** which is not yet a DiSSCo member could contribute significantly to the CANATHIST case study and benefit of the developed infrastructure to manage the collection and contribute to the GBIF and DiSSCo infrastructures.

- The **UMons** manages a smaller collection of entomology but is involved with the Africa Museum in a FedTwin profile about the pollinator insects in sub-saharan Africa. It is therefore an important partner for the prioritization of the digitization of the collections.

The 3 African partners are the main institutions related to biodiversity in their countries:

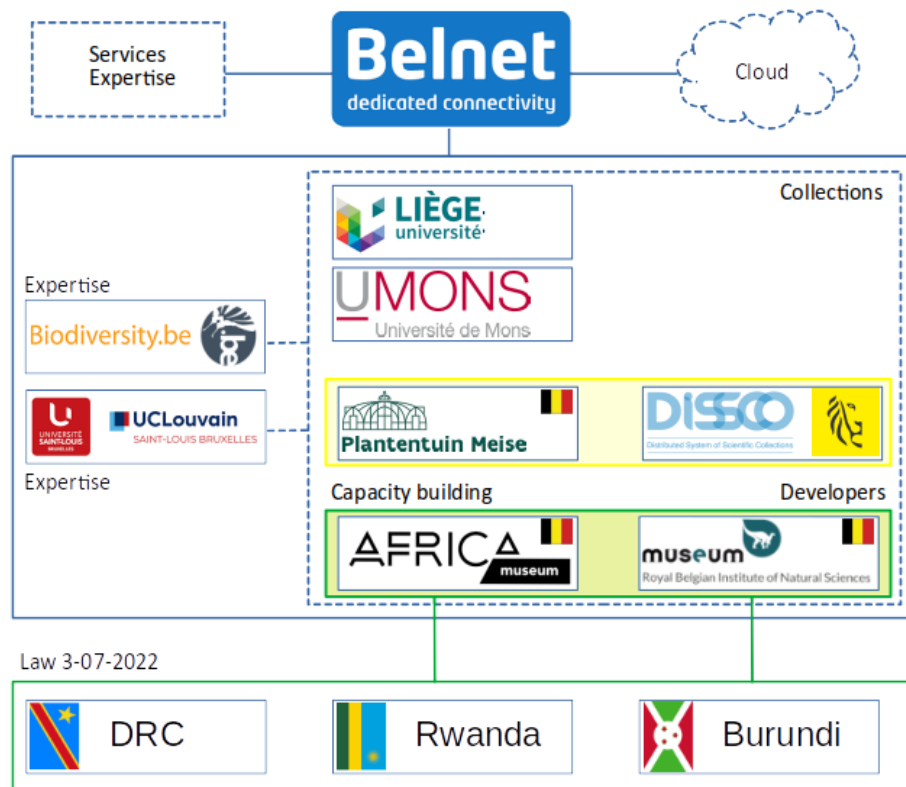
- The Centre de surveillance de la biodiversité de l'Université de Kisangani (**CSB**) aims to be the pole of excellence and expertise on the sustainable management of biodiversity in the Congo basin. It was created in 2014 with the support of the AfricaMuseum, the RBINS and the MBG and is funded by the Congolese government, the Belgian cooperation and BELSPO.
- The **CoEB** (Center of Excellence in Biodiversity and Natural Resource Management) is a Rwandan Government institution that functions as a consortium of governmental and non-governmental organizations, bringing together expertise and skills to produce information needed for effective policy and science-driven economic transformation. The Center is hosted at University of Rwanda and works in three areas: research and monitoring, bioprospecting, and conservation education and awareness raising.
- The Burundian Office for the Protection of the Environment (**OBPE**) is in charge of the application of the CBD and the Nagoya protocol in Burundi. It coordinates the protected areas and the inventories of biodiversity. OBPE is also in charge of the detection and control of invasive species as well as pest control.

#### 4.1.2 Collaborative approach

The collaborative approach is based on previous bilateral collaborations related to the different aspects of the proposal:

- **RMCA** and **RBINS** have collaborated for several years in the digitization of their collections. The collaboration of the 2 FSIs is structural. They already share technical platforms, common database systems and digitization staff and manage together the Belgian federal zoological collection. Both FSIs collaborate also with **MBG** (which holds the 3rd federal natural history collection size wise in Belgium) in the framework of Belgian and international projects (e.g. NaturalHeritage, DiSSCo, Synthesys+). People know what facilitates communication between partners and common work. This long term collaboration will facilitate the setup of the concerned platform. MBG federates Flemish scientific institutions and also ULB and UNamur from the French community in the framework of [DiSSCo Flanders](#).
- The **ULiège** and the **UMons** are not yet members of DiSSCo. ULiège manages the largest collection of entomology of the Belgian French speaking community. UMons and Africa Museum collaborate already in the framework of a FedTwin project on the pollinators of important crops in Sub-Saharan Africa.
- The **USaintLouis** already collaborates with Africa Museum and RBINS in the framework of the [BRAIN 2.0 HOME](#) project.
- The **Biodiversity platform** collaborates with all Belgian partners in the preparation of datasets for the GBIF data infrastructure.
- **Belnet** provides internet access to all Belgian partners and hosts services/tools like the Bibliography ORFEO portal or the FileSender service for large files transfer.

- **RMCA, RBINS** and **MBG** collaborate with the **RDC, Rwanda** and **Burundi** partners in the framework of training and capacity building on taxonomy, collection management, and biodiversity inventories funded by the Belgian Development Agency. The African Partners will also participate in the prioritization of the digitization activities defining the criteria and the aims of the digitization as a reference for the training in taxonomy, the biodiversity inventory and the sustainable exploitation of natural resources.



The CANATHIST consortium composed by 7 Belgian partners and 3 African partners

The CANATHIST proposal will federate the interests and the competences of the Belgian and African partners. Together they will build the infrastructure in order to provide an optimal access to this unique heritage to the different stakeholders:

- RMCA and RBINS will develop the different blocks of the CANATHIST infrastructure using open source technologies and use cases provided by the Belgian and African partners.
- The proposal will build on regular e-meetings which demonstrated all their potential during the COVID 19 crisis. This approach is safer because it is sometimes difficult to obtain visas for African participants and has the best ratio cost/efficiency.
- The partners will also meet physically in the framework of international meetings like UN Biodiversity Conferences (COP) and in the framework of the training and capacity building organized in Central Africa.
- Ad hoc physical meetings between Belgian partners will be organized for specific topics (e.g. training on digitization techniques, data validation, ...).

#### 4.1.3. Follow-up Committee & Stakeholders' involvement

Applicants are encouraged to fill out the  *Follow-up Committee letter of intent* even though the document is not compulsory.

The Follow-up committee will be composed of experts and stakeholders relevant to the project's scope. They are representatives from institutions / working groups not included in the project consortium, and experts from different stakeholders with a key role at the Belgian level, such as from the Belgian Biodiversity Platform and members of Belgian ABS Focal point.

During the project lifetime the members of the Follow-up committee will be kept informed of the regular progress of the project and they will be granted access to the project website. They will be asked to test the outputs and give feedback on the adequateness of it for their expectations and needs. They will be invited to participate in the project's official meetings and training events. Whenever relevant also asked to participate in ad hoc meetings (e-meetings or via the forums).

Ana Casino	Executive Director of the CETAF (European Consortium of Taxonomic facilities) and Deputy Coordinator of DiSSCo Coordination and Support Office (CSO) (acasino@ceta.org)
<u>Henk Vanstappen</u>	specialist in digital collections and linked (open) data, Datable (henk@datable.be)
Dr Salima Kempnaer	ABS Belgian Focal point, General Directorate (DG5) Environment
Dr Hendrik Segers	Coordinator National Focal Point Biodiversity Convention
<u>Dr Quentin Groom</u>	Secretary of the Biodiversity Information Standards (TDWG)
Dr Eric Laureys	Open data and Open Access directive, Scientific and Technical Information Service
<u>Dr Roald Hayen</u>	Responsible Laboratories Department of the Royal Institute of Cultural Heritage
<u>Dr Serge Lemaître</u>	Curator of the America Collections of the Royal Museums of Arts and History, Brussels

The underlined names have already explicitly expressed their (specific) interest in the project through a letter of intent.

#### 4.1.4 Gender Equality Plan

The primary criteria for selection of staff (paid or not of project funds) are their competences, their willingness, commitment and availability to contribute to the project. But the partners of the project endorse and are committed to respect equality and to perform staff appointment with no form of positive or negative discrimination based on gender, social or ethnic origin, religion beliefs, sexual orientation, age or disability; as elaborated in the Belgian Scientific Institutions and Universities ([VLIR Charter: Gender in Academia](#)) statements. The gender issue is not directly addressed by the project but the virtual access to information without the need of physical travelling can be evaluated as a positive aspect for tackling gender discrimination as young parents hesitate to travel for long periods.

## 4.2 DETAILED DESCRIPTION OF THE WORKPLAN

### 4.2.1 Detailed description of the work plan

Please replicate the table as many times as necessary, adding WP2, WP3...

<b>WP 1: Coordination, Project Management and Networking</b> (Month 1 – Month 36)	
<i>RMCA (Leader) (1.5 PM) (in-kind 2.1 PM), RBINS (1.5 PM) (in-kind 1 PM), MBG (in-kind 0.1 PM), USaintLouis (in-kind 0.2 PM), ULiège (in-kind 0.1 PM), UMons (in-kind 0.1 PM), Belgian Biodiversity Platform (in-kind 0.1 PM), OBPE (in-kind 0.1 PM), CSB (in-kind 0.1 PM), CoEB (in-kind 0.1 PM)</i>	
The goal of this WP is to coordinate the implementation of the project as outlined in the work plan; ensure progress of the project, the adherence of the partners to administrative guidelines of the program as required by BELSPO and so guarantee the timely delivery of the project's outputs. Furthermore this WP will ensure a proper communication flow between the funding agency, the partners of the project, the follow-up committee and with the international contacts.	
Task 1.1 Project Coordination (Month 1 – Month 36)	
<i>RMCA (Leader) (in-kind 0.5 PM)</i>	
Project lead, chairing the Project Follow up Committee; identification and troubleshooting of organizational problems. Liaison between BELSPO services and the project partners. Ensure effective communication of administrative tasks and issues within the partners. Representation of the project. Production and consolidation of cost-statements	
Task 1.2 Networking (Month 1 – Month 36)	
<i>RMCA (Leader) (in-kind 0.3 PM), RBINS (in-kind 0.3 PM)</i>	
This task will ensure proper communication and coordination between partners, WPs, the follow up committee and other external contacts. It will also organise as needed the project assembly meetings (can be in the form of e-meetings). It will put in place the project collaborative platform (using a Plone server already installed at RBINS)	
D.1.1.1	Introductory meeting (Month 1)
D.1.1.2	Midterm meeting (Month 18)
D.1.1.3	Closure meeting (Month 36)
Task 1.3 Project Reporting (Month 1 – Month 36)	
<i>RMCA (Leader) (1.5 PM) (in-kind 1.3 PM), RBINS (1.5 PM) (in-kind 0.7 PM), MBG (in-kind 0.1 PM), USaintLouis (in-kind 0.2 PM), ULiège (in-kind 0.1 PM), UMons (in-kind 0.1 PM), Belgian Biodiversity Platform (in-kind 0.1 PM), OBPE (in-kind 0.1 PM), CSB (in-kind 0.1 PM), CoEB (in-kind 0.1 PM), Belnet (subcontractor)</i>	
Production and consolidation of periodic external reports. Work plan maintenance, monitoring of the work packages	
D.1.2.1	Preliminary Report of the Project (Month 6)
D.1.2.2	Intermediate Report of the project (Month 20)
D.1.2.3	Final report of the Project (Month 36)
<b>WP 2: Data Management</b> (Month 1 – Month 30)	
<i>RMCA (Leader) (in-kind 0.2 PM), RBINS (in-kind 0.5 PM)</i>	
The tasks of this work package are to carry out the follow-up of the data management plan and its publication on dmponline.be, to guarantee the backup of the data collected by the project on the institutional servers, to	

	ensure the export of the data and metadata to the international aggregators (GBIF) and to set up the long-term backup of the data and metadata on the BELSPO LTP platform.
D.2.1.1	updated version of DMP (Month 3)
D.2.1.2	saved on dmponline.be (Month 3)
D.2.1.3	Backup on the BELSPO LTP (Month 3-36)

**WP 3: Inventory of the natural history collections collected in central Africa by Belgian institutions during the colonial period and state of the art in the national context.** (Month 1 – Month 18)

**RBINS** (Leader) (11.5 PM) (in-kind 7.1 PM), **RMCA** (11 PM) (in-kind 2.2 PM), **MBG** (in-kind 1.3 PM), **ULiège** (in-kind 0.4 PM), **UMons** (in-kind 0.4 PM), **Belgian Biodiversity Platform** (in-kind 1 PM), **OBPE** (in-kind 0.2 PM), **CSB** (in-kind 0.2 PM), **CoEB** (in-kind 0.2 PM)

Task 3.1 Belgian Survey among the DiSSCo institutions and non DiSSCo scientific institutions (Month 1 – Month 36)

**RBINS** (Leader) (5 PM) (in-kind 2 PM), **RMCA** (1 PM) (in-kind 0.5 PM), **MBG** (in-kind 0.6 PM), **ULiège** (in-kind 0.2 PM), **UMons** (in-kind 0.2 PM), **Belgian Biodiversity Platform** (in-kind 0.3 PM)

The consortium will prepare a survey among the Belgian institutions establishing the list of natural history collections from Central Africa Origin

D.3.1.1	prototype of the survey (Month 3)
D.3.1.2	raw data collected with the survey (Month 6)
D.3.1.3	analysis of the data provided by the survey (Month 8)

Task 3.2 Identification of the most important collections and their provenance study (Month 9 – Month 18)

**RBINS** (Leader) (6PM) (in-kind 4.1 PM), **RMCA** (9 PM) (in-kind 0.1 PM), **MBG** (in-kind 0.5 PM), **ULiège** (in-kind 0.1 PM), **UMons** (in-kind 0.1 PM), **OBPE** (in-kind 0.1 PM), **CSB** (in-kind 0.1 PM), **CoEB** (in-kind 0.1 PM), **Belgian Biodiversity Platform** (in-kind 0.3 PM)

The consortium will prepare the list of the collections and extract the localities of provenance. It will create the most complete multilingual list of names of localities with associated geocoordinates using existing web services and gazetteers.

D.3.2.1	list of the collections derived from the D 3.1.3 (Month 9)
D.3.2.2	geographic origin of the collections (Month 12)
D.3.2.3	list of the localities for multilingual thesaurus (Month 18)

Task 3.3 Selection of the testing collections for the setup of the infrastructure (Month 12 – Month 18)

**RBINS** (Leader) (0.5PM) (in-kind 1 PM), **RMCA** (1 PM) (in-kind 0.3 PM), **MBG** (in-kind 0.2 PM), **ULiège** (in-kind 0.1 PM), **UMons** (in-kind 0.1 PM), **Belgian Biodiversity Platform** (in-kind 0.4 PM), **OBPE** (in-kind 0.1 PM), **CSB** (in-kind 0.1 PM), **CoEB** (in-kind 0.1 PM)

This task will evaluate the collection using the priority criteria of Africa partners, the criteria of non-funded partners and produce the prioritization of the digitization efforts. CeBios will organize workshops with the African partners to define their criteria.

D.3.3.1	priority criteria defined by African partners (Month 12)
D.3.3.2	priority criteria defined by non-funded partners (Month 12 )
D.3.3.3	list of the collections used as case studies for the setup (Month 18)

**WP 4: Legal background and work regulations (USaint-Louis, all)** (Month 1 – Month 18)

**USaintLouis** (Leader) (in-kind 11.6 PM), **RMCA** (in-kind 1.4 PM), **RBINS** (in-kind 1.7 PM), **OBPE** (in-kind 0.7 PM), **CSB** (in-kind 0.7 PM), **CoEB** (in-kind 0.7 PM), **Belgian Biodiversity Platform** (in-kind 1PM)

The work package will evaluate the exact impact of the CBD convention and the Nagoya ABS protocol on the digitization program of the Belgian Natural History scientific institutions as the scope of the Nagoya protocol is



still debated between the north and the south countries. The digitization of reference material allowing biodiversity inventories is therefore a point of attention which has to be addressed in this project. The work package will also analyse the complementarity and/or the conflict of interest between the [Nagoya protocol](#) and the [Open Access directive](#) entered into force on 16 July 2019 and transposed in national law on 16 July 2021.

Furthermore, the work package will analyse broader heritage legislation that also applies to natural history collections and specifically rules concerning their digitisation and restitution.

Finally, some ethical analysis - crossed with legal sources - will be conducted in order to address questions surrounding data ownership, sovereignty and access, in particular for the data made available on the proposed platform.

Task 4.1 Legal framework of the Natural sciences collection (Month 1 – Month 12)

**USaintLouis** (Leader) (in-kind 11 PM), **RMCA** (in-kind 0.4 PM), **RBINS** (in-kind 0.3 PM), **Belgian Biodiversity Platform** (in-kind 1PM)

This task will defines the Belgian and international legal framework related to the Natural History collection

D.4.1.1 Nagoya protocol and Access Benefit Sharing (Month 3)

D.4.1.2 Open Access directive and Open Data (Month 6)

D.4.1.3 Heritage legislation and digitisation (Month 9)

D.4.1.4 Ethical and legal questions on data ownership, sovereignty and access (Month 11)

Task 4.2 Impact of the CBD and Nagoya protocol (Month 12 – Month 15)

**USaintLouis** (Leader) (in-kind 0.6 PM), **RMCA** (in-kind 0.3 PM), **RBINS** (in-kind 0.3 PM), **OBPE** (in-kind 0.2 PM), **CSB** (in-kind 0.2 PM), **CoEB** (in-kind 0.2 PM)

The Task 4.2 will evaluate the exact impact of the CBD convention and the Nagoya ABS protocol on the digitization program of the Belgian Natural History scientific institutions as the exact scope of the Nagoya protocol is still debated between the north and the south countries. The digitization of reference material allowing biodiversity inventories is therefore a point of attention which has to be addressed in this project.

The task will also analyze the complementarity and/or the conflict of interest between the [Nagoya protocol](#) and the [Open Access directive](#) entered into force on 16 July 2019 and transposed in national law by 16 July 2021.

D.4.2.1 Impact for RDC (Month 15)

D.4.2.2 Impact for Rwanda (Month 15)

D.4.2.3 Impact for Burundi (Month 15)

Task 4.3 Establishing MOUs between collections institutions and African partners (Month 15 – Month 18)

**RMCA** (Leader) (in-kind 0.7 PM), **RBINS** (in-kind 1.1 PM), **OBPE** (in-kind 0.5 PM), **CSB** (in-kind 0.5 PM), **CoEB** (in-kind 0.5 PM)

RMCA and RBINS will prepare specific MOUs with the OBPE (Burundi), CSB (DRC) and CoEB (Rwanda)

D.4.3.1 MoU RDC (Month 18)

D.4.3.2 MoU Rwanda (Month 18 )

D.4.3.3 MoU Burundi (Month 18)

**WP 5 : development of technical infrastructure (WP leaders: RMCA & RBINS)** (Month 6 – Month 30)

**RMCA** (Leader) (31,5PM) (in-kind 1.7 PM), **RBINS** (19 PM) (in-kind 12 PM), **Belgian Biodiversity Platform** (in-kind 2 PM)

This WP consists of analyzing and developing a suite of open source tools for FSIs and other Belgian institutions to manage and valorize their digital collections, and to identify the standards and formats needed to import and extract data, and for communication between the components of the CANATHIST stack.

Task 5.1 technical analysis (6 PM) (Month 1 – Month 12)

**RMCA** (Leader) (3,5PM), **RBINS** (in-kind 1 PM), **Belgian Biodiversity Platform** (in-kind 1 PM)



Task 5.1 will elaborate the architecture of the databases, webs services and decentralized tools (for data input and reporting) and synchronization strategy to be developed within the project. The focus will be put on accounting real user requirements (especially those of Central African scientists) and to the usability of well-known and mature data standards that are already used by a wide community. The analysis will be driven by the definition of case-studies assessing how scientists check and correct the most critical parts of an ecological data model in terms of data quality issues (e.g. taxonomic duplicate, handling of synonyms and vernacular names, variation in geographical place names and coordinates formats...) and develop a semi-automated workflow to speed up the correction of data. Possibility to store several versions and trace the history of a dataset, both for desktop and server-side databases will be investigated.

The project will not dig into APIs for semantic web that require a complex server-side infrastructure. Still, data quality will be guaranteed both by the data model and by the conception of the graphical interface for data input that will be linked both to external and an internal semi-opened controlled vocabularies, documenting the provenance of terms and keywords.

D.5.1.1	Documentation of Case studies (Month 6)
D.5.1.2	Reference document for Sources for base data and controlled vocabularies (Month 6)
D.5.1.3	Reference document documenting the existing data standards and software solution to be used within the project (Month 6)
D.5.1.4	Specifications and reference document for the architecture, data model and communication workflows of the CANATHIST server-side applications (Month 9)
D.5.1.5	Specification and reference document for the architecture and data model desktop app (viewer /editor) (Month 9)
D.5.1.6	Strategy for the replication/synchronization of databases between the components of the CANATHIST task (Month 12)

Task 5.2 deployment/integration/enhancement of existing services (Month 6 – Month 12)

**RMCA (Leader) (5PM) (in-kind 0.5PM), RBINS (1pm) (in-kind 2 PM)**

Task 5.2 will package together the existing server side services and data repositories, following the analysis (task 5.1). Development effort will be put on adding an interoperability to these components (probably HTTP APIs in JSON formats) to make them interoperable with the rest of the CANATHIST application stack.

DarWIN will be modified in order to ease and automate its installation on a small scale data center or a local computer. The data model will also be updated to trace the version of a dataset (currently partially implemented as a log table, but mostly available to the administrator) with limited performance downturn.

The current mechanism linking Darwin records to stable (non deletable) identifiers will be generalized as a specific application able to link several data repositories and use several standards (Handle, CETAF stable identifiers, PURL, Wikidata links ). Possibility to automate backup and transfer of data into a LTPservice will be explored.

The CMS will be linked to images in IIIF format and multimedia servers. CollectiveAccess could be a solution as it offers the possibility to import and export data in CSV formats, as well as a customizable search interface with faceted search, and is compliant with ElasticSearch. The project will follow the ongoing discussions for a multimedia standard by the W3C (e.g. [W3C Media capabilities](#)). CMS will be adapted to automatically synchronize with the IIIF server, minimizing the use of intermediate data files.

D.5.2.1	adaptation of collection Management systems (packaging and historial layer) (Month 9)
D.5.2.2	Permanent identifier services (Month 12)
D.5.2.3	Image services (Month 15)
D.5.2.4	Multimedia services (Month 15)

Task 5.3 Bibliographical services (Month 12 – Month 18)

**RBINS (Leader) (2PM) (in-kind 1 PM), MRAC (4 PM) (in-kind 0 PM)**

The OAI-PMH service developed within the framework of the NaturalHeritage project (based on ElasticSearch and Symphony) will be enhanced in order to bind with several external data sources (Plone with JSON API, other OAI-PMH servers). It will allow connecting the biblio.naturalsciences.be to the ORFEO infrastructure. The Linkage with DOI and PID (permanent identifiers) and cross linking with associated documents (PDF scans,

records of type specimens cited in the article) will also be established. Importation of reference lists in BibTex format will be implemented. The proposal will also investigate how to integrate Zotero	
D.5.3.1	<i>enhanced OAI-PMH server</i> (Month 12)
D.5.3.2	biblio.naturalsciences.be connector (Month 12)
D.5.2.3	Zotero integration (Month 18)
Task 5.4 Maps server and distribution maps (Month 12 – Month 18)	
<b>RMCA (Leader) (4 PM) (in-kind 0.2 PM), RBINS (1 PM) (in-kind 1 PM)</b>	
This task will be two folded and divided between	
1) Four server-side components for	
1.1) the publishing of on-line reference maps based on the RBINS and RMCA in OGC WMS/WFS ( <i>Web Map Server, Web Feature Server</i> ) format distribution data and background layers (biological and geological data)	
1.2) the building of interactive maps that to be embedded in web pages (based JavaScript libraries such as OpenLayers and services such as the LizMap plug-in of QGIS server) via a codeless list of parameters defying the map layers, legend and coloring scheme. The system will also be transactional, allowing modification of maps (e.g. as a transactional WFS reachable by QGIS or LizMap) and tracing historical versions of the maps.	
1.3) These maps will also be used in the query interfaces of the CANATHIST CMS (DARWiN data fitting into a specific natural reservation, watershed or administrative area).	
1.4) setting up of a <a href="#">Geonetwork</a> , OGC CS-W ( <i>Catalog Service for the Web</i> ) for geographical metadata (GUI and web-service for geographical metadata of collections, datasets and geographical maps)	
2) Two desktop components	
2.1) querying the own CANATHIST server and external resources (OpenStreetMap and OverPassTurbo API, GeoNames, Gerry Geographical Thesaurus) to georeference (as point with radius error, line and delimited polygons) or perform reverse geocoding operations (derivate place names from coordinates)	
2.2) easing the extraction of static map files from distribution data in spreadsheets formats that can be embedded in Word or LaTeX publication with predefined background (elevation data with standard color scheme, list of natural reservations, watersheds)	
D.5.4.1	server-side WMS/WFS service with user-defined interactive maps (Month 18)
D.5.4.2	client desktop application for geocoding, reverse geocoding and extraction of static map files (Month 18)
Task 5.5: Annotation server (text and image) 6PM (RBINS) + 12 PM for 3d measures from 2d images	
<b>RBINS (Leader) (15 PM) (in-kind 4 PM), RMCA (in-kind 0.5 PM)</b>	
This task will link the CANATHIST service task to existing annotation tools (e.g. annotation layer of IIIF server such as Mirador and UniversalViewer and libraries such as Labelstudio that are closer to image recognition and machine learning). Data from the CANATHIST CMS (taxonomies) will also provide controlled inputs for the annotations. The work package will add a metrological layer and will also be appended to the IIIF server and 3D viewer.	
CANATHIST will also develop a “Metrological tool” allowing scientists to extract 3D landmarks from a set of oriented 2D images (e.g. a set of focus stacking images). This approach is very promising for small specimens combining the extended view of high resolution focus stacking and the 3D morpho-geometry analysis. This is the complementary approach of the ScAnt project which proposes an Open Source low cost high resolution digitization setup.	
D.5.5.1	Annotation service (Month 12)
D.5.5.2	Metrological service on 2D images sets (Month 12)
Task 5.6: Indexation and server-side user interface	
<b>RMCA (Leader) (10 PM), RBINS (in-kind 1 PM)</b>	

The Task 5.6 is related to the aggregated indexation system used by the CANAHIST infrastructure. This approach was validated by the NaturalHeritage BRAIN project and allows to create links between data and metadata stored in different information systems.

The proposal will develop the aggregated indexes using the Elastic Search technology (community branch) or the OpenSearch (Open Source version of Elastic Search with tightened security and supported by AWS).

[Elasticsearch](#) (and/or [OpenSearch](#)) are distributed, RESTful search and analytics engines, with a focus on speed, facet visualization, and highly-available clusters with a level of fault-tolerance .

- It centrally stores data for fast search, fine-tuned relevancy, and powerful analytics.
- Elasticsearch runs to any web server, AWS S3 cloud servers with petabytes of data and laptops with Java runtime

The proposal will also evaluate and configure the [Dejavu Web UI](#) for Elasticsearch as it facilitates the import of data in Elasticsearch indexes, the creation of search views and reporting. Dejavu is an open-source MIT licensed product and can be installed using Docker.

The main task of this WP will be the definition and implementation of the technical architecture synchronizing the “master” server-side databases and APIs to the desktop tool developed in WP 5.6. This synchronization strategy must offer a good balance between performances, a limited load on the Internet connection (especially for the African CANATHIST partners) and a level of automation making it simple and transparent to the users. Several architecture will be envisioned and compared (master-slave replication, cluster of indexes) at different level (use of REST API, synchronization at database level, exchange of static files). The BELSPO services (such as Filesender and eventually the technical file standards used by the LTP) will provide tools for this task

D.5.6.1	Definition of the data model of the indexes and APIs (Month 12)
D.5.6.2	Development of graphical user interface and tools to administer, query, import and extract data from/to the index (Month 12)
D.5.6.3	Synchronization between APIs, server-side indexes and the WP 5.6 desktop tool on the user side (Month 12)

#### Task 5.7: Offline data management and local valorization of data

**RMCA** (Leader) (7PM) (in-kind 0.5 PM), **RBINS** (2 PM) (in-kind 2 PM), **Belgian Biodiversity Platform** (in-kind 1 PM)

This task will develop tools and user interfaces, aimed at desktop computers, allowing users to manage their own collection data. .

- A graphical input interface (e.g. XLS templates and forms) will ensure data quality by the means of controlled vocabularies. These vocabularies will be available without permanent internet connectivity, while being initialized and resynchronized on-demand from master indexes.
- It will also offer the possibility to associate local files (images, sound recordings, bibliographical references) to specimens / collection data, and export them in a single compressed file (e.g. ZIP or TAR including XML/JSON metadata). The format used by the BELSPO LTP could be used as a basis.
- The interface will offer a search engine working both on the user dataset and a local copy of central CANATHIST indexes (distribution data, taxonomy, collection data, bibliography) that will be refreshed from the master index when connectivity is available.

Users will be offered to navigate to online resources (specimen records in Darwin, taxonomic information in [Catalog of LIFE](#) and [GBIF](#), locality in [OpenStreetMap](#), articles with a [DOI](#) resolver).

It will also offer the possibility to export data in several formats for data exchange (DarwinCore for GBIF, DaRWIN import templates, static regional checklist in [LaTeX](#) format, [DublinCore](#) with associated images and maps, template for IIF server such as [CollectiveAccess](#)).

D.5.7.1	Development of the decentralized indexes and associated local tools for data management including controlled vocabularies (Month 18)
D.5.7.2	Development of tools to link and export data at user-side level (LaTeX, Maps, DarwinCore) (Month 18)
D.5.7.3	Development of the graphical user-interface to search and synchronize data (Month 18)

<b>WP 6: Training and capacity building (RBINS)</b> (Month 24 – Month 36)	
<i>RBINS (Leader) (4 PM) (in-kind 5.5 PM), RMCA (2 PM) (in-kind 7 PM), MGB (in-kind 0,5 PM), Belgian Biodiversity Platform (in-kind 1 PM), OBPE (in-kind 1 PM), CSB (in-kind 1 PM), CoEB (in-kind 1 PM)</i>	
The aim of this WP is to establish a framework for strengthening a partnership and collaboration between the Belgian and the Central African partners for data sharing and research capacity building. It will address a strong requirement made by the African partners to share the biological digital data about Rwandan, Burundi and RDC ecosystems and biodiversity that are hosted at RMCA, RBINS and BGM, which may include collections data, photographs, maps, archives and publications. Training will be done through different projects/programs financed by the Belgian cooperation.	
Task 6.1 Data and metadata collect of natural history collections ( <b>Biodiversity Platform</b> ) (Month 1 – Month 12)	
<i>Biodiversity Platform (Leader) (in-kind 1 PM), RBINS (2 PM) (in-kind 1 PM), RMCA (in-kind 2 PM)</i>	
This task will organize trainings and documentation related to standardization of the data and metadata of the natural history collections	
D.6.1.1	Dataset (Month 12)
Task 6.2 Low cost High Resolution Digitisation of natural history specimen (Month 13 – Month 24)	
<i>RBINS (Leader) (2 PM) (in-kind 2 PM), RMCA (2 PM) (in-kind 2 PM), MGB (in-kind 0,5 PM)</i>	
This task aims to develop locally low cost High Resolution Digitisation tools allowing scientists to digitize natural history specimens and to compare/export their data with/to the platform.	
D.6.2.1	Best practices (Month 24)
D.6.2.2	Training (Month 24)
Task 6.3 Use of the valorisation tools developed in WP 5 (Month 25 – Month 36)	
<i>RMCA (Leader) (in-kind 3 PM), RBINS (in-kind 3 PM), OBPE (in-kind 1 PM), CSB (in-kind 1 PM), CoEB (in-kind 1 PM)</i>	
This task will train the different categories of African users (e.g. IT, technicians and scientists) to use the infrastructure as active users	
D.6.3.1	Training session organized in the frame of project financed by the Belgian cooperation (Month 36)
<b>WP7: Sustainable development of web services (RMCA)</b> (Month 18 – Month 36)	
<i>RMCA (Leader) (3 PM) (in-kind 1 PM), RBINS (in-kind 2.5 PM), Belnet (subcontractor), Belgian Biodiversity Platform (in-kind 0.5 PM)</i>	
The consortium will set up the sustainability and business plan with the support of Belnet.	
Task 7.1. Proven sustainability through Proof of Concept (Month 18 – Month 36)	
<i>RMCA (Leader) (1 PM) (in-kind 0.5 PM), RBINS (in-kind 0.5 PM), Belnet (subcontractor)</i>	
Provision of a common POC infrastructure to host all services and tools during the project, and ensuring that the platform is made available to each of the partners within the project.	
D.7.1.1	POC Cloud infrastructure (Month 18)
Task 7.2 Sustainable development of the online platform and development of a business model for the web services Evaluation of the requirements to ensure technical sustainability, and how to optimize costs and use of resources to ensure sustainable development of the Belnet online platform. (Month 18 – Month 36)	
<i>RMCA (Leader) (in-kind 1 PM), RBINS (in-kind 0.5 PM), Belnet (subcontractor)</i>	
D.7.2.1	Guidelines for sustainability (Month 36)
D.7.2.2	Business model (Month 36)

	Task 7.3 Compliance of the web services with European research infrastructures Data transformation guidelines for the integration of data from the online platform into European RIs ( DiSSCo, Europeana) by providing data in specific data formats. (Month 18 – Month 36)
	<b>RBINS (Leader) (in-kind 2 PM), RMCA (2 PM), Belgian Biodiversity Platform (in-kind 0.5 PM)</b>
	This task will train the different categories of African users (e.g. IT, technicians and scientists) to use the infrastructure as active users
D.7.3.1	Services requirements for data integration in EU RIs (Month 24)
	<b>WP8: Valorisation, dissemination (RBINS) (Month 1 – Month 36)</b>
	<b>RBINS (Leader) (4 PM) (in-kind 3.5 PM), RMCA (3PM) (in-kind 3.5 PM), BGM (in-kind 0.8 PM), USaint Louis (in-kind 1.7 PM), Uliège (in-kind 0.3 PM), UMONS (in-kind 0.3 PM), Belgian Biodiversity Platform (in-kind 0.5 PM), OBPE (in-kind 0.3 PM), CSB (in-kind 0.3 PM), CoEB (in-kind 0.3 PM)</b>
	This package concerns the internal and external dissemination of the project. Communication activities will have the aim to engage all participants, related agents and interested parties. RMCA will coordinate these actions in order to reach the concerned stakeholders.
	Task 8.1. Training of the DiSSCo and non DiSSCo scientists (Month 1 – Month 36)
	<b>RBINS (Leader) (1PM) (in-kind 1 PM), RMCA (1 PM) (in-kind 1 PM), BGM (in-kind 0.5 PM), USaint Louis (in-kind 0.2 PM), Belgian Biodiversity Platform (in-kind 0.2 PM)</b>
	This includes the training of the DiSSCo and non DiSSCo scientists concerned by the development of the platform in order to test the tools implemented
D.8.1.1	Training session organized by and for DiSSCo and non DiSSCo scientists (Month 18-36)
	Task 8.2. Scientific dissemination (Month 1 – Month 36)
	<b>RMCA (Leader) (2PM) (in-kind 1.5 PM), RBINS (2 PM) (in-kind 1.5 PM), BGM (in-kind 0.3 PM), USaint Louis (in-kind 1.5 PM), Uliège (in-kind 0.3 PM), UMONS ((in-kind 0.3 PM), Belgian Biodiversity Platform (in-kind 0.3 PM), OBPE (in-kind 0.3 PM), CSB (in-kind 0.3 PM), CoEB (in-kind 0.3 PM)</b>
	The consortium will present the state of the art and the input of the CANATHIST Infrastructure to several audiences in the framework of conferences and workshops. It will also publish in scientific journals to present the CANATHIST infrastructure and the outputs of the project.
D.8.2.1	Scientific communication in workshops and conferences (Month 1-36)
D.8.2.2	Scientific papers in Impact Factor journals by the partners(Month 12-36)
	Task 8.3. Website (Month 1 – Month 36)
	<b>RBINS (Leader) (1PM) (in-kind 1 PM), RMCA (in-kind 1 PM)</b>
	The project website will be based on the Plone CMS hosting the naturalheritage.be portal
D.8.3.1	Website, CANATHIST online platform (Month 1-36)

#### 4.2.2 Work planning and time schedule: GANTT chart

Please complete the  GANTT chart and leave this section empty.

#### 4.2.3 Implementation risk management plan (max. 1,5 pages)

Number, identify and explain the main incurring risks that could delay or hinder the project and the contingency plans foreseen to deal with them, replicating the table as many times as required.

#### R1a: IT Staff not available or appointed by other projects during the project

The IT profiles for the setup of the CANATHIST infrastructure are demanding staff with skills in Open Source technologies and experience in projects. The impact is moderate, the likelihood is “Unlikely”.

- The 2 main developers of the infrastructure are already recruited on long term contracts in both FSIs. This limits the likelihood of the risk while the recruitment of new IT developers will be a high risk for the project due to the lack of IT profiles on the market.

**R1b: IT Staff not available or appointed by other projects after the project**

The CANATHIST infrastructure will need support after the end of the CANATHIST project.

- CANATHIST is a part of the core service of the IT department of the AfricaMuseum.
- 2 main developers of the infrastructure are long term contracts in both FSIs
- Business plan established with the help of Belnet will mitigate the risk for the other users of the developed infrastructure with clear SLA for external users
- Integration in the DiSSCo infrastructure

**R2: Data are not available for the setup of the infrastructure**

The data is not available and the infrastructure will stay an empty shell as the project itself is not funding the digitization of the collections. The impact is severe and the likelihood is "Very Unlikely".

- More than one million of data is available in the CMS of the Africa Museum, RBINS and Botanic Garden Meise.
- More than 70000 pictures are digitized
- Several thousands of PDFs of publications are available.

**R3: New digitization is not possible**

The setup of the CANAHIST infrastructure requires new digitization which is not funded by the project itself. The impact is significant and the likelihood is "Very Unlikely".

- Digitization is one of the priority of scientific institutions
- Digitization of the collections is supported by the DIGIT-4 program
- The setup of the DiSSCo infrastructure will boost the digitization of Natural History collections

**R4: Risk 4: Belnet not interested to host the platform at the end of the project.**

The CANAHIST infrastructure could not be directly supported by Belnet as a service for Belnet users. The impact is minor and the likelihood is "Possible".

- Belnet will help partners to build the business plan of the infrastructure during the project
- The RMCA Data center will support the CANAHIST at least for the African Partners
- The cloud hosting is possible without the direct support of Belnet for the Belgian partners

**R5: Technology not available**

The different technologies for the setup of the infrastructure are not available. The impact is severe but the likelihood is "Very Unlikely".

- All technologies are already developed and available in Open Source
- The Metrology module is already in development by a Master thesis and the CANAHIST PM are mainly devoted to new features and integration in the infrastructure

**R6: Systems are not compatible**

The different blocks are not compatible and the integration is not possible. The impact is significant but the likelihood is "Unlikely".

- The CANAHIST project will use open standards and web service protocols
- Docker will be used for the building and sharing of containerized applications and microservices. This will help in the sustainability of the infrastructure as the blocks are freezed in a working configuration.

**R7: No political support for the infrastructure for provenance studies**

The Provenance studies are not supported by the political level and no funding is available for infrastructure related to these studies.

- The Law of 3 July 2022 is promoting the provenance studies in the processes of restitution
- It is a digitization strategy of the collections of the FSIs

#### R8: Change in the licensing scheme of one software

The licensing scheme of one of the blocks makes the infrastructure in danger. The impact is moderate and the likelihood is “Unlikely”.

- The CANAHIST will build only on Open Source Blocks or Community versions
- The project will evaluate several options and choose the best option in term of continuity of service

#### R9: Maintenance cost too high after the project

After the project, the limited budgets of the FSIs will compromise the sustainability of the CANAHIST infrastructure. The impact is moderate and the likelihood is “likely”.

- The use of open source technologies limits the license fees
- The hosting of the digitized data and the FAIR access is an obligation for FSIs thanks to the Open Data directive
- The setup of the DiSSCo infrastructure will help FSIs to propose the CANATHIST as an e-service for the DiSSCo community and propose the infrastructure as a center of excellence for the Central Africa

Locate the number of the risk (R1, R2, R3...) on the table ‘Risk likelihood vs. impact’.

		IMPACT				
		Negligible	Minor	Moderate	Significant	Severe
LI K E L I H O O D	Very likely					
	Likely			R9		
	Possible		R4			
	Unlikely			R8		R1a
	Very Unlikely			R1b	R3 R6	R5 R2

	Low
	Low-Medium
	Medium
	Medium-High
	Severe

### 4.3 DATA MANAGEMENT PLAN

#### 4.3.1. Data Management Plan

Please complete the [Data Management Plan](#) and leave this section empty.

### 4.4 BUDGET

#### 4.4.1. Budget table

Please complete the [Budget table](#) and leave this section empty.

#### 4.4.2 Justification of the requested budget



The total budget of CANATHIST is 739.211 € from which 70% is allocated to the staff. This is justified by the extent of

- IT development which aims to interconnect open source tools and adapt them to both Belgian and African users. (24 PM)
- The development of the metrology tools for 3D measures on the base of the several 2D images
- The collect and the validation of the data and metadata related to the Central Africa collections and especially the creation of a multilingual thesaurus related to the localities with associated coordinates.

The equipment expenditures are related to the acquisition of one server corresponding to the “small datacenter configuration” and one workstation related to the “offline configuration”.

25% of the budget is allocated to the subcontractor which will have to establish the sustainability of the platform over the long term and provide cloud space during the duration of the project. Belnet will also provide several tools/services such as the ORFEO, or investigate future integrations such as the possibility to interconnect FileSender with the BELPO LTP or create a PID server.

## 5. IMPACT, ADDED VALUE & SUSTAINABILITY

### 5.1. POTENTIAL IMPACT OF THE PROJECT

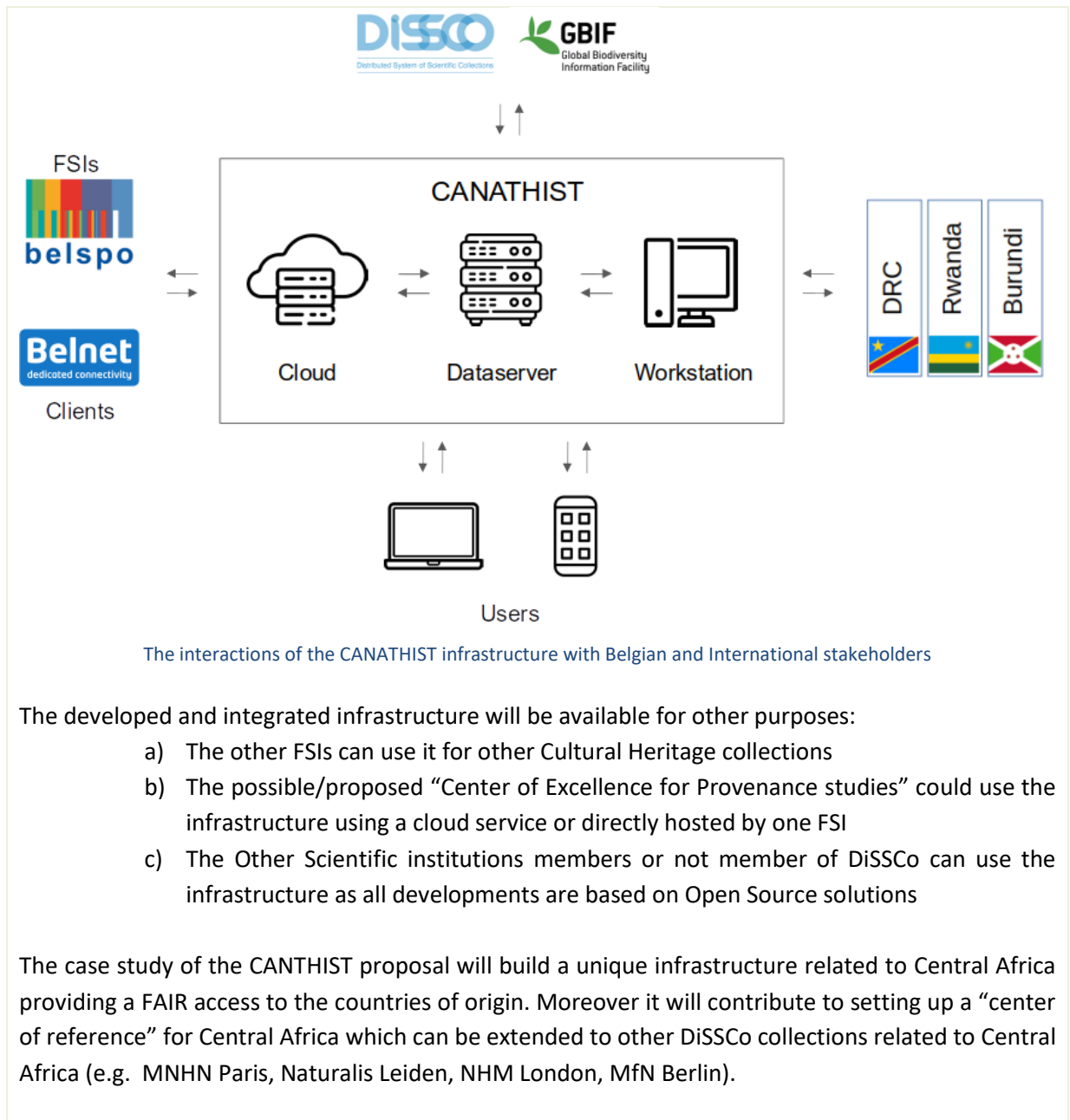
#### 5.1.1 Potential impact of the project in light of the expected outcomes and added value for the RI and the Federal Research Institution (max. 1,5 pages)

Royal Museum of Central Africa and Royal Belgian Institute of Natural Sciences are Federal Scientific Institutions with research and collections components. The specificity of the FSI's compared to Universities is the Heritage value of their collections and their mission to acquire, take care and valorise collections in an optimal way. The outputs of the CANATHIST proposal will provide to the NH FSIs the integrated infrastructure to manage and share their digital collections including the digital specimens and all associated multimedia files. The size of the Natural History collection (> 48 millions) means that the developed infrastructure will also be suitable for the collections of the other FSIs which are smaller than 1 million items. The CANATHIST proposal will therefore consider an alternate Open Source Collection Management system suitable for other Cultural Heritage (e.g. Collective Access).

The different blocks of the infrastructure will be available through the Github repository. The consortium will also prepare a business plan with the support of Belnet, for the institutions willing to benefit from the cloud version proposed through Belnet.

The small data center and the workstation versions will be more suitable for African partner institutions as the high speed internet coverage is still limited in these countries or expensive for small entities.





## 5.2. SUSTAINABILITY

### 5.2.1 Business plan

The sustainability of the developed web services - the CANATHIST online platform for access to the open source tools/services - is of crucial importance, and therefore a partnership is established with Belnet, the service provider for the Belgian research community. To test the web services, and tools developed by the partners in the project and to see if it can be sustainable, Belnet will provide space on Belnet infrastructure where all services and tools can be hosted for a proof of concept (POC) platform (*Task 7.1*). Belnet will also investigate with the project partners to see what is needed and what revenues to market the solution are necessary to ensure the technical sustainability of the platform into the future (*Task 7.2*).

In order not to hinder the already foreseen future developments, the exploration of data transformation possibilities to align the web services generated by CANATHIST with European Research Infrastructures (e.g. DiSSCo) (*Task 7.3*) have been mentioned.

Belnet will help to keep the platform sustainable on the long term defining

- In-kind contribution(s) of the institution(s) and how it is likely to change over time.
- External stable sources of (co-)funding, including users fees.
- Estimated operation time, network activity plan.
- Suitability model (location of the RI)

This will provide a sustainability model for the infrastructure which will help the AfricaMuseum and RBINS to maintain the CANATHIST infrastructure in the long term.

The CANATHIST infrastructure will be integrated in the Collection Services of the 2 FSIs and contribute with the support of the Belgian Biodiversity Platform and the services of Belnet to the e-services proposed by the Belgian federal level to the DiSSCo infrastructure.

Sharing all information, digital specimens and archives with the competent authorities, institutions and citizen communities in both Central Africa and Belgium will ensure long-term valorization of the data. The long term preservation of the data, metadata and associated multimedia related to CANATHIST will be achieved using the Belspo Long Term Preservation infrastructure.

The digitization of the collections (Belspo DIGIT program) will continue to feed the CANATHIST infrastructure after the end of the project, while the long term activities of the RBINS Cebios will continue to provide training and capacity building to the Africa partners after the project itself.