



**HAL**  
open science

## The e-COL+ Project, an Opportunity to Reflect on the Concept of Digital Twin

Pauline Provini, Jonathan Blettery, Delphine Brabant, Franck Bellugeon, Eva Perez Pimparé, Killian Leblanc, Evariste Monvoisin, Emmanuel Robert, Emmanuel Fara, Pierre-Yves Gagnier

### ► To cite this version:

Pauline Provini, Jonathan Blettery, Delphine Brabant, Franck Bellugeon, Eva Perez Pimparé, et al.. The e-COL+ Project, an Opportunity to Reflect on the Concept of Digital Twin. *Nuncius. Annali di Storia della Scienza*, 2024, 39 (3), pp.804-820. 10.1163/18253911-0390310115 . mnhn-04781139

**HAL Id: mnhn-04781139**

**<https://mnhn.hal.science/mnhn-04781139v1>**

Submitted on 13 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# The e-COL+ project, an opportunity to reflect on the concept of digital twin

## **Abstract**

The mass digitization efforts associated with the e-COL+ project, focusing on French natural history collections, provide an invaluable opportunity to reassess the need to build a comprehensive 3D digital collection to complement existing 2D digital repositories. This article explores the importance of incorporating 3D data into digital collections, highlighting its significant impact on scientific research, conservation efforts, and public outreach campaigns. Although the e-COL+ project faces practical obstacles related to digitization, it also raises important theoretical questions about the complex interplay between physical and digital objects. Here, we challenge the concepts of digital clone and digital twin that are commonly used in the construction of digital collections. The digital twin concept more accurately captures the intricate and adaptive connection between physical and digital entities that results from the precise digitization techniques used and the evolving characteristics of the specimens found in natural history collections.

## **Key words**

Digital collections, digital twin, extended specimen.

## **Main text**

### **1. Enhancing the value of data from natural history collections, the e-COL+ project**

Awareness of threats to biodiversity and gaps in our understanding, as well as the development of new technologies, have led to a renewed interest in natural history collections. These collections, scattered in different institutions, are a valuable source of knowledge, enriched over 350 years. They have two main roles: to serve as traditional references and as a vast repository of observations. In this sense, they are essential for the global inventory of biodiversity, the modelling of ecological change and other scientific fields. France's national natural science collections contain around 100 million specimens, almost 6% of the world total. These repositories of biodiversity and geodiversity now constitute an irreplaceable heritage and an indisputable resource for current and future scientific research.

However, because of their cumulative nature, natural history collections bring together thousands of objects, collected at different times and by different methods. By their very nature, these objects vary in size, composition, fragility and rarity, and therefore in the medium in which they are conserved. As a result, providing access to every object in natural history collections is an inextricable challenge. Yet, in an open science framework, natural history collections should be available to the widest possible audience. In recent decades, large-scale digitisation campaigns have changed the game. By compiling numerical data (especially photographs) of

specimens held in natural history collections in a comprehensive approach, their visibility, access and indexing have taken a spectacular leap forward. Within this framework, the Réseau national des collections naturalistes<sup>1</sup> (Récolnat) is a research infrastructure (RI) that brings together French natural history collections and experts to promote natural history collections as tools for science culture, and society. The overall mission of the Récolnat infrastructure is to produce, mobilize, aggregate, harmonize and provide access to naturalist collections and their metadata (Pérez and Pignal, 2013). The aim is to provide the scientific community with high-quality natural history data, at the right scale, interoperable and in line with international standards, to enable them to respond to major environmental and societal challenges. The infrastructure provides access to a portal containing almost 11 million images and their metadata, from more than 80 institutions in France (metropolitan and overseas). This corpus of data is also enriched by the participatory science programme (e.g., Les Herbonautes<sup>2</sup>, whose community - with over 4,000 contributors - builds up datasets for selected research projects (Heaton and Millerand, 2018)). The infrastructure also provides online tools for researchers working on digital duplicates of the collections (Annotate-On<sup>3</sup>). Led by the Muséum national d'Histoire naturelle (MNHN), Récolnat is also the French hub of the European mirror

---

<sup>1</sup> <https://www.recolnat.org/fr/>

<sup>2</sup> <http://lesherbonautes.mnhn.fr/>

<sup>3</sup> <https://www.recolnat.org/en/annotate>

infrastructure DiSSCo<sup>4</sup> (Distributed Systems of Scientific Collections (Hardisty et al., 2020)).

To date, most of the digital objects available are images (in two dimensions, 2D). Building on Récolnat, the e-COL+ project aims to digitise in three dimensions (3D) a wide range of complete or sub-complete specimens, representative of the taxonomic diversity of extant and extinct metazoans held in French natural history collections.

This project is part of a French program (Structuring Equipment for Research (ESR / EquipEx +) supported by the French National Research Agency (ANR-21-ESRE-0053<sup>5</sup>). It is led by the MNHN and involves 9 other French research institutions<sup>6</sup>. Starting in June 2021, it will run for 8 years, divided into two phases: 4 years of implementation, 4 years of massive digitization. This project will enable the French research institutions involved to set up mass digitization workflows by: 1) acquiring advanced equipment adapted to the digitization of natural collection objects, covering a wide range of sizes, textures and conservation methods (dry, liquid, pinned material, etc.); 2)

---

<sup>4</sup> <https://www.dissco.eu/fr-fr/>

<sup>5</sup> <https://anr.fr/ProjetIA-21-ESRE-0053>

<sup>6</sup> Institutions involved in the e-COL+ project:

[Muséum national d'Histoire naturelle](#)

[Centre Informatique National de l'Enseignement Supérieur](#)

[Centre National de la Recherche Scientifique](#)

[Conservatoire national des Arts et Métiers](#)

[École Pratique des Hautes Études](#)

[Institut de Recherche pour le Développement](#)

[Université de Bourgogne](#)

[Université Claude Bernard Lyon 1](#)

[Université de Montpellier](#)

[Sorbonne Université](#)

providing the means to develop and innovate in the optimization of digitization processes, data standards and long-term preservation by ensuring the FAIRisation of data<sup>7</sup> (Findable, Accessible, Interoperable, Reusable (Wilkinson et al., 2016)).

The scanning techniques used in the e-COL+ project fall into three main categories: Surface scanning, photogrammetry and computed tomography (CT) scanning (Waltenberger et al., 2021, and Wils' paper in this issue).

Surface scanning is used to collect a cloud of points corresponding to the geometry of the object's surface. These points, usually collected using surface scanners, can then be used to extrapolate the shape of the object by creating a polygon mesh. If colour information is collected at each point, the colours or textures of the object's surface can also be mapped onto the digital surface model.

Photogrammetry uses a least squares algorithm to reconstruct the shape, colour and texture of an object's surface from multiple images taken by cameras.

X-ray micro-tomography is used to obtain cross-sections of an object. Internal density variations are used to reconstruct a virtual 3D model that reflects the shape of the object's external and internal structures, allowing virtual dissections of the digital object.

These three techniques should be adapted to the specimen to be digitized, especially to its size and preservation medium. They generate information

---

<sup>7</sup> <https://www.go-fair.org/fair-principles/>

associated with the scanning technique and the digitized object, called metadata. At a minimum, this information includes the inventory number of the digitized object, but can also include more specific information (e.g., name of each digitized part, name of the operator who scanned the object, resolution used for scanning, etc.). The ability to link this group of information to other databases with matching entries is powerful. For example, the specimens digitized by each institution involved in the e-COL+ project are recorded in the Global Biodiversity Information Facility<sup>8</sup> (GBIF) through their collection numbers and a unique identifier. This allows information from the digitized object to be linked to additional information, such as its full taxonomy or collection location. These types of links can bring the number of database entries associated with the digitized specimen to a high level of complexity.

The e-COL+ project also aims to develop artificial intelligence (AI) tools to automate the segmentation and labelling of the different parts of the 3D models resulting from the digitisation of specimens, thus facilitating feature recognition and analysis. As the 3D digitisation of several thousand of specimens in natural science collections generates a colossal amount of data, the e-COL+ project also prompts us to reflect on data standardisation, based on the FAIR principles of Open Science (Vicente-Saez and Martinez-Fuentes, 2018). Indeed, an important goal of the e-COL+ project is to

---

<sup>8</sup> <https://www.gbif.org/>

develop tools to store and archive digitised data and to make them available easily, quickly and efficiently on a centralised online platform, freely accessible to researchers and the public.

## **2. Why is it relevant to mass digitize naturalist collections in 3D?**

### *1. Scientific interest*

Biological structures unfold in all three dimensions of space. Considering the ability of 3D digitization to provide detailed morphological information and enhanced visualization, while preserving the specimen integrity, digitizing specimens from natural history collections in all directions was a logical development.

In recent decades, 3D digitization tools have become increasingly powerful, and the MNHN has gradually acquired state-of-the-art equipment and expertise to digitize specimens from its natural history collections in 3D (e.g., Surfaçus<sup>9</sup> for surface scanning and photogrammetry, created in 2008, AST-RX<sup>10</sup> for X-ray computed tomography in 2011). Beyond pre-existing facilities, the e-COL+ project enlarged the 3D technical arsenal of the MNHN, Université Claude Bernard Lyon 1 (UCBL) and Université de Bourgogne (UB), through the acquisition of X-ray microtomographs, surface scanners and photogrammetry set-ups (e.g., analytical platform “collections and 3D imaging”; <https://lgltpc.fr/tous-plateforme/collections-de-geologie->

---

<sup>9</sup> <https://3dtheque.mnhn.fr/platform/surfacus>

<sup>10</sup> <https://3dtheque.mnhn.fr/platform/astrx>



et-imagerie-3d/).

In 2016, a complete digitization of the mounted skeletons of the Palaeontology and Comparative Anatomy Galleries was carried out (Gagnier et al., 2017). The spectacular results mainly benefited the MNHN's dissemination missions, as the data obtained from the digitization cannot be easily used for research. In fact, the specimens were digitized as a block of point clouds, which prevents each specimen from being separated from the others. The difficulty of segmenting these point clouds to extract individual 3D models hindered the ability to use them for comparative anatomical studies or other shape analysis (Gaillière, 2021).

The digitization methods used in the e-COL+ project address this limitation by using surface scanning and X-ray micro-tomography to generate surfaces of the digitized object (meshed 3D objects). This allows the individualization of anatomical structures of interest so that they can be viewed from all angles through virtual handling. Their shape can be accurately quantified, in particular by morphometric geometric methods, providing high-density shape data (Goswami et al., 2019), which provides valuable information to answer evolutionary questions (e.g., Leblanc et al., 2023). One of the advantages of X-ray tomography is the ability to rapidly digitize entire specimens, whether they are fluid preserved or frozen (Keklikoglou et al., 2021). This allows external and internal structures to be visualized in 3D, with the advantage of preserving their position within the organism and with little damage to the specimen (Kersemans et al., 2011). The connections

between structures can be studied, including the detailed anatomical complexity at the level of articulations between multiple bones or segments. Bringing to light hidden areas that would be lost in osteological preparation provides new information and helps to innovate anatomical study methods. The creation of such digital morphological resources, essential for answering new questions in evolutionary biology, is therefore closely linked to the development of tools for non-invasive, high-resolution digitization of whole specimens. Over the last decade, the development and increasing accessibility of 3D digitization tools has motivated numerous international institutions to propose projects for the 3D digitization of large numbers of specimens from natural history collections. Dozens of national and international projects aim to produce high-resolution anatomical data of most vertebrate species (oVert<sup>11</sup>), providing best practices and standards that include effective procedures, preferable resolutions, and archiving formats that maximize the range of downstream applications (Watkins-Colwell et al., 2018). Some projects focus on a particular group, with the aim of digitizing all known fish (Scan All Fish<sup>12</sup>) or snakes (Scan All Snakes, yes, all of them!<sup>13</sup>) and making these data available to the public and the research community, without any restriction. This new wealth of high-resolution anatomical data allows for comparative analyses that were previously impossible (Muñoz and Price, 2019; Provini et al., 2023). By proposing a

---

<sup>11</sup> <https://www.floridamuseum.ufl.edu/science/overt/>

<sup>12</sup> <https://osf.io/ecmz4/wiki/home/>

<sup>13</sup> <https://lsa.umich.edu/ummmz/news-events/all-news/search-news/scan-all-snakes.html>

new dataset that goes beyond the model species traditionally used, the corpus constituted by the e-COL+ project could open new research opportunities in many fields (biomimicry, comparative anatomy, systematics, etc.), while meeting conservation challenges.

## ***2. Conservation interest***

### *1. Computerization and databases updating*

Since the specimens digitized for the e-COL+ project must have an inventory number, they must all have been computerized to appear in the databases containing information on the collections. Each digitized specimen must appear in the application used to computerize the collections of each institute it belongs to (e.g., JACIM at the MNHN) and in the Récolnat database, playing the role of the database of reference. Moreover, the "encyclopaedic" approach, which calls for the creation of a corpus representing the taxonomic diversity of natural history collections, confronts us with the need to have as accurate a picture as possible of what the collections are about in order to be able to make a selection. This requires a quantitative and qualitative analysis of the taxonomic diversity on the one hand, but also of the nature of the objects that represent it (completeness of specimens, nomenclatural status, type of conservation in particular). However, collections consist of millions of specimens that are far from being inventoried and registered in databases. Considering only specimens that have already been computerized would not allow us to consider the coverage of taxonomic diversity. It was therefore

decided to work with curators to find specimens that have not yet been inventoried, and to include them in the database as they are digitized. The e-COL+ project will thus contribute to the completion of the inventories and contributes to the efforts to computerize the French natural history collections, which will help to increase the value of the specimens by improving their visibility and searchability

This encyclopaedic approach will also lead e-COL+ to work in collaboration with other research programs that aim to inventory living organisms and complete French natural history collections. One example is the Faune de France initiative run by Patrinat<sup>14</sup>, a joint support and research unit, which aims to enhance the biodiversity present in France (Dumeau et al., 2022). Some programmes, such as freshwater fish in France, could be enriched by additional data from a specific e-COL+ axis (Keith et al., 2020). This would require e-COL+ to focus part of its work on this area and provide additional data. Another example of a research programme, which may be a great opportunity of collaboration with the e-COL+ project is ATLASea<sup>15</sup>, which plans to collect and interpret the DNA of 5,000 marine species from the French coasts. Again, e-COL+ could add species targeted by ATLASea to its corpus.

By working with other research programmes, e-COL+ is becoming part of

---

<sup>14</sup> <https://www.patrinat.fr/fr>

<sup>15</sup> <https://jacob.cea.fr/drf/ifrancoisjacob/Pages/Actualites/Vie-Institut/2022/ATLASea-laureat-PEPR-equipement-recherche.aspx>

the ongoing exploration of the concept of extended specimens, linking together a variety of metadata (collection, taxonomic, DNA, 2D-3D imaging, etc.).

The development of IT tools dedicated to the e-COL+ project is fully integrated into the overhaul of the MNHN's and Récolnat information system (overhaul of the 3Dthèque<sup>16</sup>, Colhelper<sup>17</sup>, JACIM, SI Récolnat<sup>18</sup>) and will eventually be linked to European infrastructures (e.g., DiSSCo).

## *2. Collection enhancement*

The e-COL+ project is an integral part of preservation missions of the MNHN, the UCBL and the UB. Some collections, especially those in liquid media, require special attention. This is the case for some of the bird collections, where the liquid medium in which they are preserved has partially evaporated. Reconstituting the liquid could damage specimens by modifying the current chemical balance, but keeping specimens semi-submerged is not a permanent solution (Herbin et al., 2020). Digitizing them in 3D allows us to preserve traces of these specimens.

In addition, building a 3D model of a specimen can be used to reconstruct damaged parts of a specimen or to fill in missing parts.

Although 3D digitization does not always offer the same level of detail as

---

<sup>16</sup> <https://3dtheque.mnhn.fr/>

<sup>17</sup> <http://colhelper.mnhn.fr/requests?segments=>

<sup>18</sup> <https://www.recolnat.org/fr/>

traditional casting, 3D models can also be used to design 3D printing 'casts' when the original specimens are too fragile. For example, the extremely fine structure of the fossilized wings of a *Meganeura* fossil (*Meganeura monyi*, MNHN.F.R51142, holotype) made the object particularly delicate and traditional casting unthinkable for this extinct dragonfly-like insect. It was therefore scanned and 3D printed.

### ***3. Outreach interest***

The e-COL+ project aims to make 3D digitization data available through a dedicated platform, accessible to all.

The data produced will attract the attention of the widest possible audience, as they will be visually appealing and easy to understand, thanks to a set of free dedicated visualization software (e.g., MeshLab (Cignoni et al., 2011), Blender (Soni et al., 2023)). They can therefore be used even more easily for dissemination or educational projects, as they are made available without restrictions.

In education, 3D modelling is used to study the evolution and mechanisms of living organisms (e.g., the study of jaws and teeth for feeding, the shape of the pelvis for locomotion or reproduction, etc.). This aligns with current 3D data outreach projects, such as the “Cabinet de Curiosités 3D<sup>19</sup>” developed by the MNHN, aiming at displaying 3D models of specimens of the MNHN’s collections.

---

<sup>19</sup> <https://cabinetdecuoriosites3d.mnhn.fr/fr>

### **3. Reflections about the digital object created**

At different stages of the e-COL+ project, a number of questions may arise, some of which remain unanswered.

#### *1. Which specimens should be digitized, which ones should not?*

The e-COL+ project has set itself the ambitious goal of producing 40,000 3D digitisations of species representative of the taxonomic diversity of fossil and extant metazoans. This diversity of species is selected from specimens inventoried in French natural history collections, in particular those of the MNHN, the UCBL and the UB. The spirit of the project is to provide an extensive taxonomic coverage of the diversity of French naturalist collections, but defining which specimens should be digitized (digitization corpus) is a complex task, taking into account multiple factors.

The MNHN's e-COL+ team focuses on digitizing a variety of complete and sub-complete specimens, representative of vertebrate and arthropod taxonomic diversity, thus we must select, among all the specimens present in the collections, complete specimens in good condition and well identified.

We prioritize the digitization of type specimens, which are used as references in systematic biology, and we prioritize specimens belonging to the "Faune de France", which catalogues all species found in France.

Occasionally, we also digitize emblematic specimens, such as the killer

whale recently found in the Seine River<sup>20</sup> (*Orcinus orca*, ZM\_AC-2023-11). We also take advantage of the reorganization of collections (e.g., large, mounted bird specimens moving from the comparison room to the taxidermy shelter) to digitize specimens that need to be moved. When possible, we take the opportunity to scan specimens that come to the museum for special events, such as the Amazon jaguar (*Panthera onca*, ZM\_2023-2) during its visit to Paris for the Félin exhibition<sup>21</sup>.

The UCBL and the UB have adopted a dual approach: the digitization of fossils and living taxa representative of their natural history collections (mammals, especially carnivorans, primates, rodents, perissodactyls, cetartiodactyls; but also, birds, crocodylians, molluscs, arthropods and trace fossils). These institutions digitize their type specimens in relation to the Récolnat database. The project also prioritizes figured specimens, as well as complete and partial specimens of rare taxa or those that are exceptionally well preserved as fossils. They are expanding this corpus to include specimens used in current research projects of the different e-COL+ teams. In addition, 3D modelling of other regional collections beyond the walls of the participating institutions has begun.

Since different equipment is used to digitize specimens, depending on their size and type of conservation, we create lists of specimens with the same type

---

<sup>20</sup> [https://www.lemonde.fr/planete/article/2022/05/29/1-orque-en-difficulte-dans-la-seine-va-etre-euthanasiee\\_6128110\\_3244.html](https://www.lemonde.fr/planete/article/2022/05/29/1-orque-en-difficulte-dans-la-seine-va-etre-euthanasiee_6128110_3244.html)

<sup>21</sup> <https://www.mnhn.fr/fr/expo-felins>



of conservation, of equivalent size and, if possible, located near the appropriate digitizing equipment, to optimize logistics and digitization time. Finally, to manage the digitization data for dissemination purposes and for linking to existing databases (Récolnat, DiSSCo), the digitized specimen must have an inventory number linked to the institution where the collections are housed.

The digitization intentions are coordinated with all the partner institutions and international digitization programs (e.g., Scan All Fish<sup>22</sup>) to avoid digitizing the same species more than once.

Despite the broad guidelines set-up in the e-COL+ project to develop the digitization corpus, the selection of specimens to form a coherent corpus is an arduous task that raises conceptual questions about the selection of specimens: if two complete specimens of the same species are available, why digitize one rather than the other? The variety of selection criteria does not always allow us to easily decide, making coordination and collective decisions even more important.

## ***2. What is the relationship between the digital and the physical object?***

In essence, the digital model created is not the same as the digitized physical object.

First, what is found in the box of a particular collection specimen is not

---

<sup>22</sup> <https://osf.io/ecmz4/wiki/home/>

necessarily intended to be digitized and included in the e-COL+ corpus. For example, a disarticulated skeleton stored in a box usually comes with crumbs and scraps. These items are part of the box where the specimen is stored, but are not likely to be digitized because of their small size and identification issues. What is the size limit that distinguishes a crumb from an actual anatomically relevant part?

Second, the resolution chosen to digitize an object depends not only on the device used, but also on the size of the object or its level of complexity. For example, some high-resolution scanning tools cannot be used for large objects; there is a trade-off between resolution and digitized data size that guides the operator in selecting the appropriate resolution or equipment. In all cases, even with the appropriate tools and resolution levels, some details are inevitably lost during digitization. By definition, surface scanning does not provide the 3D shape of the internal structures of the sample and hidden parts are not always accessible for scanning. This is the case with mounted bird skeletons, where the interconnected bones hide much of the joints or for which the shape of the rib cage are difficult to access with a surface scanner, making it impossible to obtain 3D models of complete bones with surface scanning is (Figure 1). These scanning imperfections may be acceptable, especially if no technical solution can be found to overcome them. In the case of mounted bird skeletons, X-ray CT scanning is technically difficult due to their fragility, the presence of metal parts in the mounts, and the size of some

specimens that are incompatible with the tools available in the e-COL+ project. Therefore, although it does not faithfully reflect the entire shape of the bird, surface digitization was chosen for these specimens.

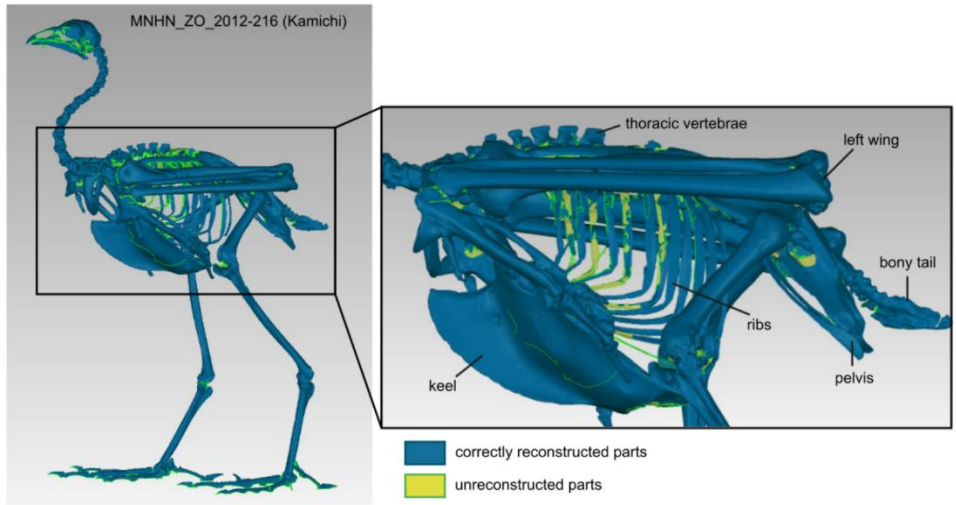


Figure 1: 3D reconstruction of a Kamichi specimen (MNHN-ZO-2012-216), created using surface scanning data acquired with Faro Edge Arm and the scanner laser blue line probe. The blue areas correspond to parts that were successfully reconstructed, while the yellow areas, outlined by green lines, represent areas that could not be reconstructed due to insufficient data.

In addition, it is sometimes difficult to determine the criteria for judging whether a scan is acceptable or not. For example, if 95% of a digital model is faithful to the shape of the object, but 5% of the object has insufficient resolution to distinguish certain structures, can the scan be considered acceptable? This depends on the type of structure for which the scan is less faithful to the shape of the physical object. For example, the mouth parts of insects are used for systematic identification and play an important role in

the essential function of ingesting food, and are located on the animal's head, a particularly symbolic structure (Figure 2). Another example is the wings of insects, which are particularly difficult to digitize and yet particularly important to understand the ecology of the animal (Schubnel et al., 2023) or for taxonomy. If the resolution is not sufficient for these parts of the specimen, it seems logical not to keep the scan data, even though the rest of the body is correctly scanned. Determining which structures need to be correctly digitized and at what resolution level thus implies a certain hierarchy of object's parts based on criteria that are difficult to objectify and requiring a high level of expertise.

The level of acceptance also depends on the target audience. If the end purpose is morphological study, the resolution must be high enough to ensure good quality measurements of the structure of interest. If the end goal is education, especially in biodiversity, a large number of complete scans at relatively low resolution from a variety of species will be more valuable than a few scans at high resolution. These potential tensions or conflicts between public interest and scientific interest further complicate the choice of specimens and digitization techniques.

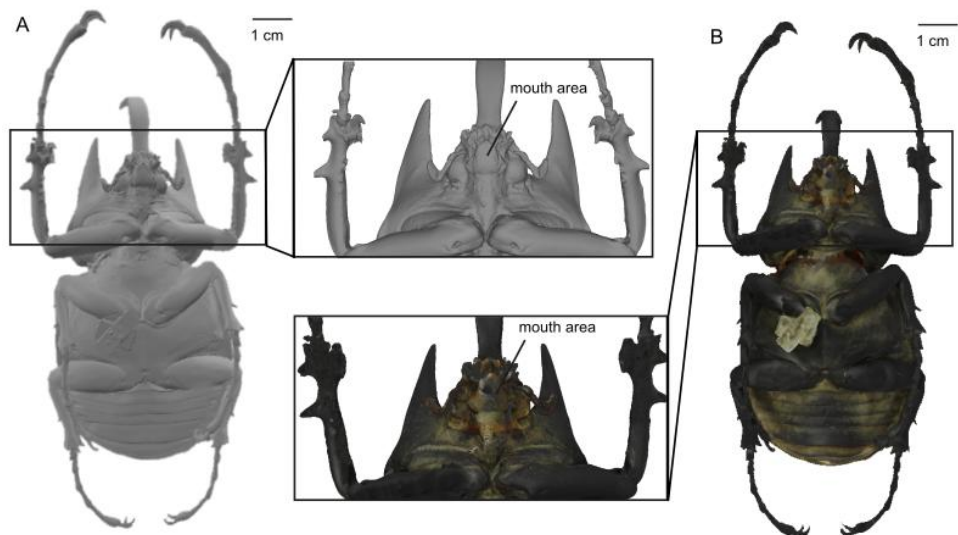


Figure 2: 3D Model of a specimen of *Megasoma action*, created using surface scanning data acquired with Faro Edge Arm and the scanner laser blue line probe, without texture (A) and with texture added to the mesh, using photogrammetry (B). The rectangular boxes show a zoom on the head area, highlighting the mouth area.

Third, a collection object and its associated data evolve over time. Conservation is now seen as a complex practice with retroactive effects on objects and subjects (Hölling, 2017). It aims not only to preserve objects, but also to understand how their identity and meaning are influenced by various factors such as time, environment, values, politics, economics, conventions, culture, and the transmission of skills, tradition, memory, and tacit knowledge (Hölling, 2017). Specimens change as they are moved, studied, collected, etc. by scientists or for collection management purposes. Therefore, the digitized model reflects the shape of the physical object at a particular point in time. The object may deteriorate over time (e.g., due to fungus, bacteria, poorly controlled sealing).

In the context of the e-COL+ project, should we digitize the same object several times if its shape changes drastically? If so, according to what criteria?

In addition, large-scale digitization brings to light errors in the identification of bones or specimens. This raises several questions. Should the digitization accurately reflect what's in the box corresponding to the inventory number? Should errors be corrected, even if it means creating differences between the physical object and the digital one? Sometimes there are errors that cannot be corrected. For example, creating a list of complete and partial specimens in the MNHN's collections revealed boxes containing bones from multiple specimens of seals. Because it is impossible to determine which bone(s) belong to the original specimen of the box, the obvious error remains unresolved.

Although the digital object reflects an "instant" in the specimen's life, it is important to consider the possibility of correcting or modifying it after the fact. Or should we opt for a digital object frozen in time?

Conceptually, the aim of large-scale digitization of naturalist collections is to produce a digital image of the objects in the collection. The goal would be to obtain a digital clone, faithfully corresponding to the entire physical collection. However, as shown here for natural history objects, by essence, the process of digitization will never provide a perfect clone of the physical

object. This partly comes from decision process on data collection, which can influence how reality is represented (Korenhof et al., 2023). Therefore, compared to the notion of “clone”, the notion of “twin” better captures the theoretical framework associated with the complexity of digitizing large-scale naturalist collections. Interestingly, this distinction is also being made in other fields today, notably in IT, engineering sciences and health (e.g., Amram et al., 2023). Identical twins are close in form but possibly slightly different since the beginning of their lives, highlighting the initial inherent difference between the physical and digital object due to the digitization process. Later in life, twins may live closely together or completely apart, reflecting the fact that the physical object may change over time, whereas the numerical object corresponds to an image at a given time that is unlikely to change in the long term. In addition, like human twins, which are separate entities, digital and physical objects can be viewed separately. In the case of the digital object available through a platform, it is possible to observe and interact with it and its metadata from a location other than where it is stored. However, the development of the digital twin, as promoted by the e-COL+ project, does not mean the death of the physical specimen and will not prevent researchers from coming to study it in person, in the collections under the care of the curators.

## **Conclusion**

The mass digitization of French natural history collections associated with

the e-COL+ project is an opportunity to reflect on the relevance of building a 3D digital collection to complement the 2D digital collections that already exist. The added value of using 3D data is evident in terms of scientific, conservation and public outreach interests. The very concrete, practical issues we face in the e-COL+ project reflect theoretical questions related to the relationship between physical and digital objects. Among these, the notion of the digital clone, commonly invoked when building a digital collection, seems less relevant to us than the notion of digital twin, which reflects the looser link that exists between the physical and digital object due to the digitization techniques used and the intrinsic life of the objects present in natural history collections. This notion also captures the idea that like a naturalist's specimen, an occurrence of the living in time and space, the digital object represents an occurrence of the physical object at a given moment, which is not fixed in time.



## References

- Amram, B., Klempler, U., Leibler, Y. and Greenbaum, D.** (2023). In Their Own Image: Ethical Implications of the Rise of Digital Twins/Clones/Simulacra in Healthcare. *Am. J. Bioeth.* **23**, 79–81.
- Cignoni, P., Ranzuglia, G., Callieri, M., Corsini, M., Ganovelli, F., Pietroni, N. and Tarini, M.** (2011). MeshLab.
- Dumeau, B., Fauvel, T., Garrido, M., Grall, J., Ouisse, V., Pasqualini, V., Pergent-Martini, C., Pergent, G., Rollet, C. and Ruitton, S.** (2022). PATRINAT.
- Gagnier, P.-Y., Maschner, H., Gailliègue, A., Norgeot, L., Dapogny, C., Révéret, L. and Abourachid, A.** (2017). Automatic reconstruction of polygon triangulation for mounted skeleton point cloud. In *2017 IEEE 13th International Conference on e-Science (e-Science)*, pp. 528–532. IEEE.
- Gailliègue, A.** (2021). Reconstruction de surfaces à partir de nuages de points d'objets complexes 3D: application à des squelettes de vertébrés complets et montés.
- Goswami, A., Watanabe, A., Felice, R. N., Bardua, C., Fabre, A.-C. and Polly, P. D.** (2019). High-Density Morphometric Analysis of Shape and Integration: The Good, the Bad, and the Not-Really-a-Problem. *Integr. Comp. Biol.* **59**, 669–683.
- Hardisty, A., Saarenmaa, H., Casino, A., Dillen, M., Gördderz, K., Groom, Q., Hardy, H., Koureas, D., Nieva De La Hidalga, A. and Paul, D. L.** (2020). Conceptual design blueprint for the DiSSCo digitization infrastructure-DELIVERABLE D8. 1. *Res. Ideas Outcomes* **6**,.
- Heaton, L. and Millerand, F.** (2018). La manifestation de l'expertise sur une plateforme collaborative du Web: le cas de l'herbier numérique Les Herbonautes.
- Herbin, M., Santin, M. D., Toro, R., Heuer, K., Cuisin, J., Martin, C., Raveendran, L. and Debruyne, R.** (2020). Do not Dispose of Historic Fluid Collections: Evaluating Research Potential and Range of Use. *Collect. Forum* **34**, 157–169.
- Hölling, H.** (2017). The technique of conservation: on realms of theory and cultures of practice\*. *J. Inst. Conserv.* **40**, 87–96.
- Keith, P., Poulet, N., Denys, G., Changeux, T., Feunteun, E. and Persat,**

- H. (2020). *Les poissons d'eau douce de France*. Biotope éditions.
- Keklikoglou, K., Arvanitidis, C., Chatzigeorgiou, G., Chatzinikolaou, E., Karagiannidis, E., Koletsa, T., Magoulas, A., Makris, K., Mavrothalassitis, G., Papanagnou, E.-D., et al.** (2021). Micro-CT for Biological and Biomedical Studies: A Comparison of Imaging Techniques. *J. Imaging* **7**, 172.
- Kersemans, V., Thompson, J., Cornelissen, B., Woodcock, M., Allen, P. D., Buls, N., Muschel, R. J., Hill, M. A. and Smart, S. C.** (2011). Micro-CT for anatomic referencing in PET and SPECT: radiation dose, biologic damage, and image quality. *J. Nucl. Med.* **52**, 1827–1833.
- Korenhof, P., Giesbers, E. and Sanderse, J.** (2023). Contextualizing realism: An analysis of acts of seeing and recording in Digital Twin datafication. *Big Data Soc.* **10**, 20539517231155061.
- Leblanc, K., Pintore, R., Galvão, A., Heitz, E. and Provini, P.** (2023). Foot adaptation to climbing in ovenbirds and woodcreepers (Furnariida). *J. Anat.* **242**, 607–626.
- Muñoz, M. M. and Price, S. A.** (2019). The Future is Bright for Evolutionary Morphology and Biomechanics in the Era of Big Data. *Integr. Comp. Biol.* **59**, 599–603.
- Pérez, E. and Pignal, M.** (2013). Numériser et promouvoir les collections d'histoire naturelle.
- Provini, P., Camp, A. L. and Crandell, K. E.** (2023). Emerging biological insights enabled by high-resolution 3D motion data: promises, perspectives and pitfalls. *J. Exp. Biol.* **226**, jeb245138.
- Schubnel, T., Mazurier, A., Nel, A., Grandcolas, P., Desutter-Grandcolas, L., Legendre, F. and Garrouste, R.** (2023). Flat does not mean 2D: Using X-ray microtomography to study insect wings in 3D as a model for comparative studies. *Methods Ecol. Evol.* **14**, 2036–2048.
- Soni, L., Kaur, A. and Sharma, A.** (2023). A Review on Different Versions and Interfaces of Blender Software. In *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 882–887. IEEE.
- Vicente-Saez, R. and Martinez-Fuentes, C.** (2018). Open Science now: A systematic literature review for an integrated definition. *J. Bus. Res.* **88**, 428–436.

**Waltenberger, L., Rebay-Salisbury, K. and Mitteroecker, P.** (2021). Three-dimensional surface scanning methods in osteology: A topographical and geometric morphometric comparison. *Am. J. Phys. Anthropol.* **174**, 846–858.

**Watkins-Colwell, G., Love, K., Randall, Z., Boyer, D., Winchester, J., Stanley, E. and Blackburn, D.** (2018). The walking dead: status report, data workflow and best practices of the overt thematic collections network. *Biodivers. Inf. Sci. Stand.*

**Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B. and Bourne, P. E.** (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9.

### **Captions for illustrations**

Figure 1: 3D reconstruction of a Kamichi specimen (MNHN\_ZO\_2012-216), created using surface scanning data acquired with Faro Edge Arm and the scanner laser blue line probe. The blue areas correspond to parts that were successfully reconstructed, while the yellow areas, outlined by green lines, represent areas that could not be reconstructed due to insufficient data.

Figure 2: 3D Model of a specimen of *Megasoma action*, created using surface scanning data acquired with Faro Edge Arm and the scanner laser blue line probe, without texture (A) and with texture added to the mesh, using photogrammetry (B). The rectangular boxes show a zoom on the head area, highlighting the mouth area.