Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
S. Barzaghi, A. Bordignon, B. Gualandi, S. Peroni – *Enlightening the Black Box of Humanities Research: Methodological Documentation as a Way to Transparency and Accountability of Digital Exhibitions*
DOI: http://doi.org/10.6092/issn.2532-8816/21177

Enlightening the Black Box of Humanities Research: Methodological Documentation as a Way to Transparency and Accountability of Digital Exhibitions

Sebastian Barzaghi

Department of Cultural Heritage, University of Bologna, Via Degli Ariani, 1, Ravenna, Italy
sebastian.barzaghi2@unibo.it

Alice Bordignon

Department of Classical Philology and Italian Studies, University of Bologna, Via Zamboni, 32, Bologna, Italy
alice.bordignon2@unibo.it

Bianca Gualandi

Department of Classical Philology and Italian Studies, University of Bologna, Via Zamboni, 32, Bologna, Italy
bianca.gualandi4@unibo.it

Silvio Peroni

Department of Classical Philology and Italian Studies, University of Bologna, Via Zamboni, 32, Bologna, Italy
silvio.peroni@unibo.it

Abstract

This article discusses open and reproducible research on Cultural Heritage by presenting the digital twin of the exhibition "The Other Renaissance: Ulisse Aldrovandi and the Wonders of the World" as a case study. After an overview of "reproducibility", "replicability" and “repeatability” in different research contexts it describes the acquisition and digitisation process, with a focus on the collection of metadata and paradata (e.g., acquisition techniques, individuals responsible, start and completion dates) and on the activities necessary to process, transform, and publish the digital cultural heritage objects and their metadata as FAIR and linked open data.

We also describe how open technologies and software were employed throughout to maximise transparency, accountability and the workflow’s re-adoption for the creation of a virtual exhibition in different settings. Finally, we circle back to “reproducible expertise” and mention the obstacles that still exist to transparency and accountability in the scholarly ecosystem.

Keywords: Transparent Research -- Open Science -- Cultural Heritage -- Digital Twin -- Reproducibility

Questo articolo affronta il tema della ricerca aperta e riproducibile nel campo del Patrimonio Culturale, presentando come caso di studio il gemello digitale della mostra “L’Altro Rinascimento: Ulisse Aldrovandi e le Meraviglie del Mondo”. Dopo una panoramica sui concetti di “riproducibilità”, “replicabilità” e “ripetibilità” in diversi contesti di ricerca, viene descritto il processo di acquisizione e digitalizzazione, con particolare attenzione alla raccolta di metadati e paradati (ad esempio, tecniche di acquisizione, responsabilità di ciascuno, date di inizio e fine) e alle attività necessarie per elaborare, trasformare e pubblicare gli oggetti digitali del patrimonio culturale e i relativi metadati come FAIR e linked open data.

*Viene inoltre illustrato come siano state impiegate tecnologie e software open source per massimizzare la trasparenza, la tracciabilità e la riutilizzabilità del flusso di lavoro nella creazione di una mostra virtuale in contesti diversi. Infine, tornando al concetto di “competenza riproducibile” (“*reproducible expertise*”), si menzionano gli ostacoli che ancora persistono alla trasparenza e all’*accountability *nell’ecosistema accademico.*

Parole chiave: Trasparenza della Ricerca -- Scienza Aperta -- Patrimonio Culturale -- Gemello Digitale -- Riproducibilità

1. Introduction

We wish to discuss open and reproducible research on Cultural Heritage by presenting a case study, the creation of the digital twin of the exhibition “The Other Renaissance: Ulisse Aldrovandi and the Wonders of the World”[1]. The creation of this digital twin has been carried out within the Project CHANGES (“Cultural Heritage Active Innovation For Next-Gen Sustainable Society”) and specifically its Spoke 4, dedicated to investigating the use of virtual technologies for the promotion, preservation, exploitation and enhancement of cultural heritage in museums and art collections [2].

The original exhibition was held between December 2022 and May 2023 in the Poggi Palace Museum in Bologna, Italy, and consisted of a collection of more than 200 objects mostly belonging to the naturalist Ulisse Aldrovandi (1522-1605) and preserved by the University of Bologna. With its large set of different small/medium objects (from manuscripts and maps to woodcuts, statues, animal models, minerals, fossils and more), the exhibition has provided an ideal experimental ground to define approaches and methods relating to the acquisition, processing, optimisation, metadata inclusion and online publication of 3D assets [2].

[1] https://site.unibo.it/aldrovandi500/en/mostra-l-altro-rinascimento

The present article extends a prior work on this topic presented at the past AIUCD Conference 2024 [4] by extending extensively the reasoning and discussion, thus providing more insights into the study conducted.

Section 2 offers an overview on the meaning of "reproducibility", "replicability" and "repeatability" in different research contexts and how they are not possible without transparency, or the careful and complete documentation of all relevant aspects of a research study.

Section 3 describes the acquisition and digitisation process, with a focus on the collection of metadata and paradata (e.g., acquisition techniques, individuals responsible, start and completion dates) and on the following software-based activities necessary to process, transform, and publish the digital cultural heritage objects and their metadata as FAIR and linked open data.

Section 4 starts from the definition of a digital replica as an approximate, aesthetically convincing copy of a cultural site or artefact [11] to delve deeper into the approach for creating the digital twin of Aldrovandi's exhibition. It also describes how open technologies and software were employed throughout to maximise transparency, accountability and the workflow's re-adoption for the creation of a virtual exhibition in different settings.

Section 5 circles back to reproducible expertise and discusses the obstacles that still exist to transparency and accountability in the scholarly ecosystem.

2. Interpreting reproducibility: current uptake

The terms "reproducibility", "replicability", and "repeatability" have historically been understood and defined in different ways across disciplines. Goodman and colleagues [17], for example, writing in the Science Translational Medicine journal, have suggested that it is possible to talk about three types of "reproducibility": (i) methods reproducibility, i.e. the ability to exactly reproduce a study by using the same raw data and the same methodologies to obtain the same results, (ii) results reproducibility – also referred to as replicability – i.e. the ability to obtain the same results from an independent study using the same methodologies as the original study, and (iii) inferential reproducibility, i.e. "the drawing of qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study". In explaining how the latter differs from the two previous categories, the authors add that scientists might "draw the same conclusions from different sets of studies and data or could draw different conclusions from the same original data, sometimes even if they agree on the analytical results" [17]. The reasons can be a priori, such as a different assessment of the probability of the hypothesis being explored or can be linked to different choices about how to analyse and report data. This third type of reproducibility is also the most important according to the authors. As noted by Leonelli [22], many articles discussing reproducible or replicable research do not usually discuss what expressions like "the same results" or "the same outcomes" refer to. Is it raw data, data models extracted from them, or generalisations derived from their analysis? This shortcoming may be traced back to an exclusive focus on experimental research that yields numerical results and a lack of consideration for other approaches to scientific research [22].

Peels and Bouter [25] have looked at how the concepts of "reproducibility", "replicability" and "repeatability" can be applied to the humanities. They prefer the terms "replicability" and "replication", and they also define three different levels that do not however overlap with those we have seen before: (i) reanalysis, that can be associated with Goodman et al.'s methods reproducibility, (ii) direct replication, where the same study protocol is applied to new data, and (iii) conceptual replication, where research data are new and the study protocol is modified [25].

The authors find that, while replication can take various forms across the humanities, it is not fundamentally different from replication in the biomedical, natural, and social sciences and can be achieved by pre-registering the studies, and documenting and sharing methodologies and data [25]. Thanks to funding from the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek or NWO), a group of Dutch researchers conducted replication studies in various fields, including some humanities disciplines (e.g., history)[2], and recently published a set of recommendations and lessons learned [12]. They found that, in all cases, replication studies help corroborate the findings of the original studies (e.g., extending the number of sources or using a more state-of-the-art approach) and can provide a more thorough understanding of the relevant research field and the available methodological choices [12]. They also note, however, that "even experienced, highly conscientious researchers often find it difficult to document their protocols in enough detail to support direct replication" [12].

Avoiding any dichotomy between humanities and so-called "hard" sciences, Leonelli [22] identifies six kinds of reproducibility that can be placed on a spectrum from computational reproducibility all the way to irreproducible research. Each kind is shaped by at least four aspects: the varying precision of research goals and control over research conditions, a different dependence on statistics and on researchers' judgement. (i) Computational reproducibility, perhaps most common in computer science and related fields, is not particularly interested in the circumstances of data production but is based on the prediction that a given input, run through the same algorithms, will produce the same output. Full replication should be possible if code and input data are made available alongside a publication. In experimental settings, such as clinical trials, variation is unavoidable. Here, (ii) Direct experimental reproducibility is obtained when there is a strong similarity between the results of different experiments: the same research methods are applied, and as much control as possible is exerted on environmental variables (albeit the degree of control is very much context-dependent). Other types of experiments are much less controllable and do not aim for direct reproducibility, as variation may be the most interesting aspect to study (it can be the case for psychology or neuroscience). Leonelli here talks about (iii) scoping reproducibility (also indirect or hypothetical reproducibility), where reliability in research can be established through the convergence of different lines of inquiry. When the control over experimental conditions is extremely limited (e.g., research on new phenomena, rare materials, unique specimens or materials that cannot be repeatedly investigated), the focus shifts to (iv) reproducible expertise. What is reproducible here are the researchers' "skills and interpretative abilities" and much importance is placed on the transparency and reliability of the methodologies employed to work on materials that are no longer accessible. Not dissimilarly, Leonelli points at fields that rely on observational techniques rather than experimentation (e.g., much research in medical, historical and social sciences) and thus put emphasis on (v) reproducible observation. Some types of ethnography, structured interviewing, and diagnoses based on medical imaging are just some examples of specific, reproducible ways of observing and studying phenomena. Finally, some disciplines reject the idea of reproducibility altogether and instead embrace the subjectivity and context-dependence of research outcomes as unavoidable. Reflexivity, transparency, and the practice of documenting, managing and preserving data and methodologies take on an even more important role in this case [22].

Carefully documenting the original study design, data collection and analysis, and reflecting on all possible influencing factors is fundamental for reliability and rigour but does not automatically ensure the reproducibility of research ([22], [26]). Those objecting to the blanket application of

[2] https://www.nwo.nl/en/researchprogrammes/replication-studies

the reproducibility or replicability categories across scientific domains note how it could potentially lead funders and evaluators to disregard the idiosyncratic and local aspects of research, and to underreport the significance of variation, with great epistemic risks [22]. Indeed, to require reproducibility of all epistemic cultures is harmful, imposes "universal policies that fail to account for local (epistemic) differences" and ultimately denies authority – and related rewards – to qualitative fields of research ([22], [26]). On the other hand, the "umbrella of Open Science" is wide enough and its "accountability toolbox" is big enough to develop plural methods for assessing the quality of diverse research practices [26]. Indeed, within the Open Science movement, researchers and advocates are trying to shift the focus of funders and evaluators on research methodologies, their discussion and documentation, and on what can be learned from unexpected and incongruent findings [22] such as negative results.

Well-defined practices are essential for ensuring transparency, reliability, and equitable access to research outcomes. In the case of the digital twin of "The Other Renaissance" exhibition, we looked for some operational indications on how to achieve this goal in the existing literature. White et al. [31] offer nine simple strategies to enhance data sharing and reuse, making it more understandable, accessible, and easier to analyse, benefitting both individual researchers working alone or together with their teams, and the broader scientific community. These guidelines, grounded in ecology and evolutionary biology, provide an introduction to good data practices that can be applied across various scientific fields, and include: (1) sharing data to enhance reproducibility and meta-analyses of results, while also providing recognition to data providers; (2) using clear and well-documented metadata to describe data in a way that makes them understandable, usable, shareable and accessible in the long term; (3) publishing data in both their raw and processed forms, along with metadata and processing code, to offer the most flexibility for users; (4) using standard, non-proprietary data formats, including clear and consistent file and table structures, to ensure that data can be easily accessed, shared, and analysed across different software and systems; (5) using consistent, software-compatible null values, to avoid confusion or errors during analysis; (6) making data easily combinable by including structured data (such as taxonomies) shared with other datasets; (7) performing quality control checks on data, to ensure its accuracy and consistency; (8) depositing data in reputable repositories for long term preservation; and (9) including explicit, open licenses with data to clearly define their usage rights, responsibilities and restrictions.

Still embracing this all-encompassing approach to transparent research, in a paper published a few years later, Wilson et al. [33] outline a set of recommendations for scientific computing, applicable across different disciplines and at varying levels of computational expertise. Regarding data management practices, their suggestions focus on the importance of incremental documentation and data cleaning. In particular, they advocate for continuous retention of raw data, robust backup strategies, data manipulation for improving machine and human readability and facilitating analysis, meticulous recording of the steps used to process data, using multiple tables in a way that each record in one table is interlinked with its respective representation in another table via a unique and persistent identifier, and using repositories that issue DOIs to the various data artefacts used and produced for easy access and citation.

In the humanities, it is well known that sharing research can be challenging due to various factors, including professional concerns, restrictive licenses, and epistemological differences over how research is framed in a particular discipline (i.e., the definition of what is data, established traditions, and so on) [15].

However, overcoming these challenges can lead to the significant benefits discussed previously, as evidenced by the growing interest in accessible data management approaches within the humanities. For example, in the archaeological context, Karoune and Plomp [20] identify three distinct levels of workflow to make research activities reproducible, depending on the computational skills required to carry out such activities, from simple public access to research materials and methods all the way to executables in a container. This three-level framework offers an entry point for researchers at all skill levels and a scalable approach that, as they gain more experience in making research more transparent, it becomes possible to take further progressive steps to enhance the reproducibility of their work in future projects.

For the digital twin of the exhibition "The Other Renaissance" the available best practices in terms of research transparency, and data and methodology management were reviewed and applied systematically. In addition to the guidelines mentioned above, data were for example collected, generated and managed according to FAIR principles [32] and reuse of existing data and standards was always favoured. Another explicit project objective was making the process repeatable – in different contexts and with different exhibition objects – creating a "template" that others could reuse. The topics that we have covered in this section provide the theoretical backbone for the project's output management strategy.

In the next sections we will see how the project's objectives have been achieved in practice. Please note however that the entire workflow is described in more detail elsewhere: for an in-depth representation of each stage and the corresponding data management, please refer to Barzaghi et al. ([5], [6]) and to the forthcoming guideline by Bordignon et al. [8].

3. Making the digitisation process transparent

To ensure a solid basis for transparency and replicability, our approach closely followed the aforementioned sets of best practices, in line with the indications listed in the Data Management Plan of the project [18]. Metadata management occurs concurrently with 3D data creation (Figure 1) and is integral to a reproducible digitisation workflow. This requires: (1) assignment of unique global persistent identifiers (PIDs) to both physical and digital objects; (2) metadata describing the objects, including their PIDs; and (3) unique PIDs for the metadata records themselves. Metadata are produced and maintained at multiple levels, encompassing both the exhibited objects and the digitisation process outputs. Metadata production and management in the digitisation workflow involved creating two datasets as Google Sheet files shared between the team members: one (Object Table, or OT) for storing catalogue descriptions of the physical objects in the collection, the other (Process Table, or PT) for storing data about the digitisation process. While not ideal for complex data management, Google Sheets proved to be a practical and familiar tool that met the project's needs. It allowed team members with minimal data management experience to efficiently collect data under tight deadlines without extensive training. In addition, the platform's real-time collaboration feature enabled simultaneous work on the data, while its version control functionality ensured that all changes were tracked throughout the process. This made it an accessible solution for the current project and a tool that can be scaled by future teams handling digitisation tasks in other projects.

After defining the structure of the tables, the variables represented by their headings, and the expected representation for each value, the data were populated in parallel by the team members. On the one hand, the OT was populated with data gleaned from official museum records and preliminary notes related to the exhibition objects and thus was structured around a cataloguing description of each object (e.g. "title", "author", and so on). Where possible, controlled data values (e.g. people names, terms used for object types, etc.) were aligned with existing vocabularies (such as WikiData[3]) and authority lists (like VIAF[4] and ULAN[5]). On the other hand, the PT was populated with data inserted by the researchers during the acquisition of the objects and the creation of their 3D models and, thus, was structured around the steps involved in the overall digitisation process and their relevant attributes. As outlined in Figure 2, the Acquisition phase (step 1) focused on capturing cultural heritage objects (CHOs) and generating the raw material (hereafter referred to as RAW) required for the creation of the corresponding digital cultural heritage objects (DCHOs). Photogrammetry and Structured Light Scanner (SLS) acquisition techniques have been implemented to obtain the digital representation of each CHO. The choice of which methodology to use has been influenced by contextual factors (such as limited time and available space), and CHO physical features (materials, shape, and size). The RAW data includes 2D image datasets for photogrammetry, while individual scans (saved as .scan files) for the SLS. The information associated with the acquisition phase encompassed the unit assigned to handle a CHO, the individuals responsible for its acquisition, the technique employed to capture the CHO's RAW data, the tools utilised during the acquisition process, and the start and completion dates of the acquisition activity.

The acquisition phase was subsequently followed by a series of software-based activities (steps 2–7), involving various tools and applications to process, transform, and publish the digital versions derived from the RAW data obtained in the preceding phase. To guarantee transparency concerning the authorial decisions made throughout these phases we provided different derivative versions for each 3D model:

- Processed Raw Data (RAWp): The initial output from photogrammetry or SLS software, containing unaltered data without interpolation or geometric corrections.
- Digital Cultural Heritage Object (DCHO): A refined model produced after addressing geometry issues, filling gaps, and applying interpolation during the 3D modelling phase using computer graphics software.
- Optimised Digital Cultural Heritage Object (DCHOo): A further optimised version created to enhance performance for real-time interaction on web-based platforms.

[3] https://www.wikidata.org/

[4] https://viaf.org/

[5] https://www.getty.edu/research/tools/vocabularies/ulan/

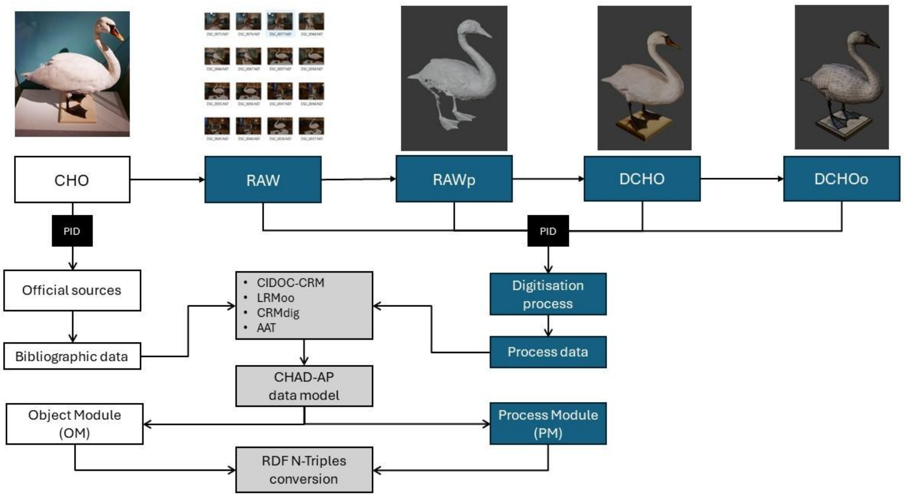

Figure 1. Parallel processes: digitisation and data management.

While the software activities may differ based on the nature of the materials being digitised and the intended purpose of the digital files, the phases we tracked in the digitisation process for all CHOs included the following:

- Processing phase (step 2): The processing of the RAW generated during the acquisition phase, specific software using automatic algorithms is used. First, when needed, adjustments are made during pre-processing using specialised editing software to correct exposure, sharpness, highlights, shadows, and white balance of the RAW. During this phase, the human operator adjusts input parameters to create the RAWp, overseeing alignment, the creation and cleaning of sparse and dense point clouds, mesh generation, and texture application.
- Modelling phase (step 3): In most cases, the complexity of the object's shape or challenging acquisition conditions prevent the complete capture of all necessary data, often resulting in models with local issues, such as gaps. In this step, the human operator addresses potential topological issues in the RAWp as part of their creative process and subjective interpretation, using computer graphics software to produce the DCHO.
- Optimisation phase (step 4): The simplification of the DCHO for specific purposes or use cases, an optimised version is created, the DCHOo. During this phase, the DCHO undergoes a retopology process to achieve a predominantly quad-based mesh. This retopology is carried out either semi-automatically or manually. The process ensures a uniform quad-dominant mesh and optimised polygon density, reducing it to an ideal range based on the features of the objects.
- Export phase (step 5): The conversion of the RAWp, DCHO and DCHOo into distinct formats. Specifically, .obj and .fbx for the RAWp and the DCHO, and .glTF for the DCHOo.
- Metadata creation phase (step 6a): The generation of e structured information of bibliographic and process data about the CHO, RAW, RAWp, DCHO, and DCHOo.

- Provenance creation phase (step 6b): Centred on developing the Metadata Record Provenance Information to monitor the agent responsible for creating bibliographic/process data, the time of creation, and the primary source of the data.
- Presentation phase (step 7): The transfer of the DCHOo from a local device or storage location to be presented on a web-based framework. Specifically, to present and publish the DCHOo, we used ATON, the open-source framework designed and developed by the CNR-ISPC in 2016 [16]. ATON leverages robust open-source platforms like Three.js and Node.js, along with solid web standards, to offer accessible solutions for organisations, researchers, and museums in developing cross-device Web3D/WebXR applications tailored to the Cultural Heritage sector.

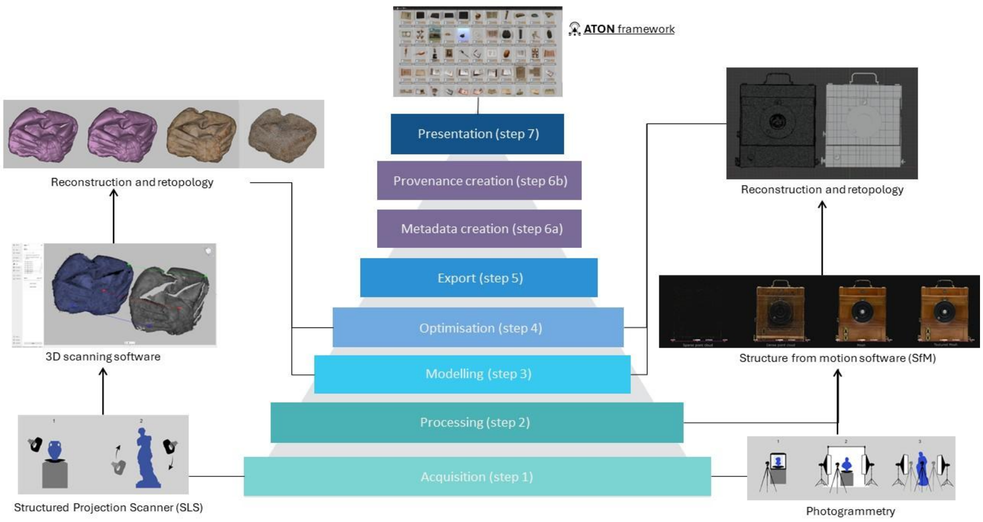

Figure 2. The acquisition and digitisation process.

For an in-depth description of each workflow stage and the corresponding data management, refer to Barzaghi et al. ([5], [6]). In this context, a dedicated guideline document is under development [8], based on practices tested within the project, to support both internal partners and external professionals involved in cultural heritage digitisation. Given the national scope of the project, the guidelines are presently available in Italian only, but we are working on producing an English version.

The creation of RAW, RAWp, DCHO, and DCHOo files, alongside their related metadata, was crucial for establishing a detailed record of the entire digitisation process. This approach allowed for effective tracking of each CHO's progress through the stages and facilitated an assessment of the project's overall success. Additionally, it ensures the long-term preservation and accessibility of DCHOs and enables a thorough analysis of the digitisation and acquisition processes, which can vary considerably in duration depending on the project's scope and the distinctive attributes of each CHO.

This preliminary work resulted in the creation of a record of the entire digitisation process. Google Sheets and Microsoft Excel in the Microsoft Office 365 platform were strong facilitators for data retention, backup and versioning[6]. Moreover, shared formatting practices on elements such as dates and names were essential for preparing the data for the subsequent phases of the project. At the end of this stage, each object had its metadata, related digitisation phases with their features, and unique identifiers that allowed the two datasets to be linked to each other. Interlinked representations of the entire physical collection, its digital counterpart, and the procedure that, from the former, produced the latter.

As information was added to both datasets, more work went into getting them ready to be published in a machine-actionable form, compliant with FAIR principles [32]. The Resource Description Framework (RDF)[7] [10] was selected as a formal data representation for enabling transparent data publishing. However, in order to transform the current data into RDF statements, the table structures had to first be mapped to data models that could express and deepen the semantics of the data about cultural heritage and digitisation activities. We chose to reuse the CIDOC Conceptual Reference Model (CIDOC-CRM)[8] [13], and some of its extensions, including the one based on the Library Reference Model (LRMoo)[9] [29] to describe the characteristics and contextual information CHOs, and CRM Digital (CRMdig)[10] [14] to depict the stages of the digitisation workflow. Moreover, the *Simplified Agile Methodology for Ontology Development* (SAMOD) [27], a process to quickly create semantic models that are supported by rich documentation and test cases, was used to draw the needed conceptual constructs from CIDOC-CRM, LRMoo and CRMdig and pack them into the *Cultural Heritage Acquisition and Digitisation Application Profile* (CHAD-AP)[11], an OWL-based application profile for describing CHOs and the processes of acquiring and digitising them into RAW, RAWp, DCHOs and DCHOos as structured, machine-actionable data [7]. CHAD-AP's structure is organised in two main modules: the Object Module (OM) and the Process Module (PM). The OM focuses on the representation of CHOs as constructs whose metadata are organised across four conceptual layers. The first layer, *Work* (`lrmoo:F1_Work`), captures the core essence of the object and is characterised by one or more titles (each an instance of `crm:E35_Title`), along with relationships to other Works. The second layer, *Expression* (`lrmoo:F2_Expression`), represents the intellectual realisation of the Work and is connected to the set of entities involved in the CHO's creation -- typically modeled as `crm:E7_Activity` and associated with elements such as `crm:E39_Actor`, `crm:E55_Type`, and others -- as well as the CHO's content, including subjects described as instances of `crm:E73_Information_Object`. The third layer, *Manifestation* (`lrmoo:F3_Manifestation`), concerns the embodiment of the Expression in a specific format and includes information about the CHO's type (`crm:E55_Type`) and its license, modelled through a semantic pattern based on `crm:E31_Document` and `crm:E73_Information_Object`. The final layer, *Item* (`lrmoo:F5_Item`), refers to the actual, physical exemplar of the CHO and is described

[6] Since transparent recording does not involve any computational code, proprietary software like Google Sheets is acceptable as long as it includes features like versioning and exporting outputs to open formats (e.g., .txt, .rtf, .pdf) [18].

[7] https://www.w3.org/TR/rdf11-concepts/

[8] http://www.cidoc-crm.org/cidoc-crm/

[9] http://iflastandards.info/ns/lrm/lrmoo/

[10] https://www.cidoc-crm.org/crmdig/; https://projects.ics.forth.gr/isl/CRMext/CRMdig_v3.2.2.rdfs

[11] https://w3id.org/dharc/ontology/chad-ap

through identifiers (`crm:E42_Identifier`) and textual descriptions, and may also include metadata related to its curation and conservation. Complementing the OM, the PM outlines the 3D digitisation workflow composed of a sequence of activities. The first of these, *acquisition* (`crmdig:D2_Digitization_Process`), engages the CHO at the Item level and produces a RAW data object (`crmdig:D9_Data_Object`). The subsequent steps, grouped as activities involving the use of specialised software (`crmdig:D10_Software_Execution`), include operations such as processing, modelling, and optimisation. Both acquisition and software activities are linked to other entities, including the 3D data (`crmdig:D9_Data_Object`), which serve as input and output chained between activities in the digitisation workflow, the individuals (`crm:E21_Person`) and organisations (`crm:E74_Group`) responsible for carrying them out, the techniques employed (`crm:E55_Type`), the tools and software used (`crmdig:D8_Digital_Device` and `crmdig:D14_Software`), and their respective temporal information (`crm:E52_Time-Span`). Both OM and PM leverage the typification mechanism introduced through `crm:E55_Type` with a limited set of ontological individuals extracted from the Getty's Art & Architecture Thesaurus (AAT)[12] [19] to specify the various entities involved at a finer-grained degree of representation.

The use of SAMOD in developing CHAD-AP facilitated the reproducibility of both the model and the development process. Exemplary scenarios of use and competency questions written in SPARQL are available on the human-readable documentation of the model[13]. The materials used for documentation and testing of the model, as well as the model itself in its multiple serialisations, are openly available on GitHub[14]. Additionally, the reuse of CIDOC-CRM and its related models as a foundational framework for the application profile aligned the semantics used to describe the CHO and the processes for creating its related RAW, RAWp, DCHO, and DCHOo to a standardised, formal language, enhancing interoperability with other realities in the cultural heritage domain.

4. Making interpretation more transparent

One of the main goals of cultural heritage digitisation is the selection of specific elements of reality to store digitally. The selection process involves a deliberate human choice about the physical, geometric, chromatic, mechanical, and stylistic characteristics of the objects to digitise. These aspects are recorded inside a "grid of information", such as vectors, images, 3D models, databases, and tables, among others [11]. According to this logic, a digital technology survey is expected to approximate reality based on some predetermined features selected at the outset of the survey project. The quantity and quality of the data obtained during the survey significantly impact how accurate the digitisation will be. In this context, a *digital replica* is defined as an approximate, aesthetically convincing copy of a cultural site or artefact [11]. In our case study, the main aim was to obtain the digital version of the exhibition's experience, starting from the

12 http://vocab.getty.edu/aat/

13 https://w3id.org/dharc/ontology/chad-ap

14 https://github.com/dharc-org/chad-ap/

creation of its digital twin[15], linking to the digital assets of the various objects (3D and multimedia) in the collections, enriched by metadata, catalogued and accessible online using different devices [2].

Our approach for creating the digital twin of Aldrovandi's exhibition included, in the first place, the implementation of various setups and instruments to create morphologically precise models with highly detailed textures. The primary challenges included the large number of CHOs to digitise (301), their diverse shapes, surfaces, and scales, as well as the limitations imposed by the exhibition's restricted time frame and physical space. Detailed documentation about these challenges and related solutions adopted in the acquisition and processing phase was created. The documentation of the risks (e.g. acquisition of non-Lambertian materials, limited object's mobility, etc.) and the solutions adopted (e.g. cross polarisation techniques, specific setup schemas, etc.) permits others to retrace and repeat, at least in theory, the actions involved in a certain research effort, producing new data [28]. Concerning SLS acquisitions, we defined some common limits regarding texture final resolution, and we decided on a specific range for geometry complexity.

During the entire process, open technologies and software were employed to maximise the workflow's re-adoption for the creation of a virtual exhibition in different settings. However, for some specific tasks (e.g. RAW elaboration), proprietary software was required since open-source software fails to produce satisfactory results.

Documenting processing decisions made for extra transparency should be a part of the scientific workflow and cultural heritage preservation. This can be done, as proposed by Moore et al. [23], by extracting a processing report from the photogrammetry software. Metashape[16] and 3DF Zephyr[17], the main software used for the photogrammetric processing phase, provide this option. The function has not been developed yet for the open-source alternative Meshroom[18], whose implementation in this project is under test. However, software for processing photogrammetric data is considered more open and transparent compared to software used for scanned data elaboration. Scanned data were elaborated using different versions of Artec Studio[19], which has proven to be a "black box" for those who do not own the software and the licence required to use it, allowing RAW export only in proprietary formats and without providing any processing report.

The challenges posed by the acquisition conditions required a blend of automated processes and manual interventions, which inevitably introduced elements of subjectivity. Preserving all 3D data derivative versions (RAW, RAWp, DCHO, DCHOo) enables direct and transparent comparison, providing a clear record of the interventions applied throughout the process.

---

[15] Since cultural heritage may be intangible or temporary, Niccolucci et al. [24] suggest separating the data exchange dimension from the representation dimension for digital twins. This reconceptualisation rethinks data flows and bi-directionality as possible and as not mandatory requirements for digital twins of cultural heritage artefacts or landscapes, opening the possibility for accurate digital models (i.e. digital replicas) to evolve dynamically into a fully developed digital twin.

[16] https://www.agisoft.com/

[17] https://www.3dflow.net/it/

[18] https://alicevision.org/

[19] https://www.artec3d.com/it/3d-software/artec-studio

Comparing the RAWp and the DCHO, for instance, enables one to identify which parts were modelled and which parts belong to the rough data elaboration. To further improve data transparency for documentation purposes, we developed a dual methodology that incorporates detailed paradata, to clearly indicate the integrated sections both in terms of geometry and texture: first, we proposed a method for documenting mesh integrations by continuously and contextually updating semantic vertex colour maps [9]; starting from that, we developed a post-hoc, method-agnostic identification approach based on false-colour texturing [1]. Finally, we used as many standard and interoperable formats as possible for the generated data to facilitate their reuse on different platforms. Respectively, we used glTF, glb, obj, and mtl for 3D models; tiff, jpg, raw, and png for images; mp4 and mov for videos; and mp3 for audio.

Specifically, we advocated for the use of glTF as the main 3D format for the DCHOo version, an open standard designed for interactive Web3D applications, ensuring high interoperability with modern 3D platforms and services, as well as facilitating the reuse and integration of licensed data within the format ([2], [30]). These decisions were documented and, in some instances, informed by the project's Data Management Plan [18].

The project's final steps are still in progress. At the time of writing, the 3D models and their accompanying data and metadata have yet to be deposited in a repository for long-term preservation. The team has chosen to use Zenodo, a general-purpose platform, as a temporary solution until a more specialized repository becomes available. While Zenodo is not tailored for 3D data or cultural heritage metadata, it assigns DOIs to deposited items, supports generic high-level metadata (DataCite Metadata Schema, Dublin Core), and is widely recognized and supported by the research community.

Zenodo was selected because it aligns with Open Science principles, is familiar to all project partners, operates independently of the institutions involved, and allows the creation of a dedicated community for the CHANGES - Spoke 4 project. This feature consolidates all project outcomes under one umbrella, avoiding fragmentation across multiple repositories. Although Zenodo is not an ideal solution for cultural heritage data, there is currently no disciplinary repository dedicated to this type of research in Italy. However, the H2IOSC project (https://www.h2iosc.cnr.it/) aims to establish a collaborative cluster of European distributed research infrastructures (CLARIN, DARIAH, E-RIHS, and OPERAS) focused on humanities and cultural heritage, with operational nodes across Italy. Once H2IOSC provides a suitable repository for cultural heritage research data, the team plans to migrate all data and metadata there.

5. Discussion and conclusions

We have described how the digitisation process of the exhibition "The Other Renaissance" has been documented throughout, with different methods, but with constant attention to research transparency, openness and accountability. Since any reality-capture or source-based model is affected by the lens of interpretation (of a human or software), tracking steps for the creation of a 3D model is essential to give transparency to these interpretations, facilitating the repeatability of the creation process [23]. Looking back at the categorisation by Leonelli [22], we are moving in the realm of reproducible expertise. We are studying unique specimens that cannot be repeatedly investigated (more on this later), the control over "experimental conditions" – e.g. the digitisation process – is limited by the constraints described in the previous sections, and the only truly reproducible element is the researchers' expertise. Again, reproducibility is made possible by the careful documentation of the methodologies and workflow employed throughout.

> Appeals to reproducible expertise are also characteristic of research on materials that are rare, unique, perishable, and/or inaccessible, such as depletable samples stored in biobanks; unique specimens, such as specific botanical finds or archaeological remains; or materials that are hard or expensive to access (such as very costly strains of transgenic mice). These materials are not amenable to repeated investigation as required by the direct and indirect forms of reproducibility. This does not constitute an obstacle to using such materials for research, since the uniqueness and irreproducibility of the materials is arguably what makes the resulting data particularly useful as evidence. The onus of reproducibility shifts instead to the credibility and skills of the investigators entrusted with handling these materials. Apposite methodologies have been developed to cope with the impossibility to directly replicate the findings, including vetted access, cross-samples research and the centralization of research in locations where several researchers can work together and check each other's work and ensure its reliability for those with no access to the same material sources. ([22], p. 137)

Data relating to the digitisation process can oftentimes be captured only once, while the process is ongoing, and it is therefore crucial to retain as much information as possible, structure it appropriately and make it available in an open and machine-readable format to provide a record of the entire physical collection, its digital counterpart, and the procedure that, from the former, produced the latter.

Furthermore, "The Other Renaissance" was a temporary exhibition. Even if it made economic sense to carry out a new digitisation campaign, applying the same (or similar) methodologies to the same cultural objects, it simply would not be possible because the physical collection has ceased to exist. Carefully documenting the research process not only allows for the same methodologies to be applied to different cultural heritage objects and exhibitions but also offers a precious additional trace of a physical collection that has been irremediably lost.

Documenting the project methodologies and workflow in this manner, although scientifically sound, is not simple: it requires careful planning, specific competencies, and it is extremely time-consuming. As noted by Peels and Bouter [25], guidelines on how to report study protocols, methodologies and procedures are needed, and this is perhaps especially true in the humanities.

In addition, there are numerous obstacles that are not intrinsic, but rather derive from how academia operates. Accountability, data curation and open, reproducible research are not normally rewarded in the academic setting.

Undoubtedly, many funding bodies have been encouraging practices such as the management of research data according to FAIR principles [32] and their publication "as open as possible". At the same time, much of the language around "open science requirements" in funding calls, including FAIR principles, often betray that same focus on experimental science and numerical results, and scarce consideration for different approaches to research that we have already mentioned with regards to reproducibility [22].

Research communities are however filling the gap, discussing and adapting FAIR principles to the different research practices ([3], [15], [21]). Data Management Plans are becoming increasingly common, and templates and online tools are being produced to help researchers plan their data-related strategies ahead of time in a structured and, possibly, machine-actionable manner. At the same time, we need to put more explicit attention on methodologies, and on the need for carefully documenting each step of a research workflow. Data, no matter how open, are of limited use if they are not sufficiently documented, and if they are not accompanied by a clear description of the research methodology.

Again, this is not always sufficient or simple – as different researchers may have a different view on data and hence require different types of contextual information – but it is certainly an important first step. To help researchers navigate the complexities of data management, Universities and research institutes are hiring new professional roles, such as data stewards, that can offer support to research groups and to institutions, offer training and draft policies and guidelines.

CoARA's *Agreement on Reforming Research Assessment*[20] defines research quality as the transparency of its processes and methodologies and as research management that allows a systematic re-use of previous results. This Coalition's commitment to reforming how researchers and their organisations are evaluated may finally give scholars the opportunity to focus less exclusively on journal (and other long form) publications and more on the planning, development, and publication of data, metadata, methodologies, and more.

6. Acknowledgements

This work has been partially funded by Project PE 0000020 CHANGES - CUP B53C22003780006, NRP Mission 4 Component 2 Investment 1.3, Funded by the European Union - NextGenerationEU.

7. Author's contributions

Authors' contribution according to CRediT (https://credit.niso.org/): Conceptualization (SB, AB, BG, SP); Investigation (SB, AB, BG); Supervision, Validation (SP); Writing – original draft, Writing – review & editing (SB, AB, BG).

8. References

[1] Ammirati, Luisa, Alice Bordignon, Federica Collina, Francesca Fabbri, Bruno Fanini, Daniele Ferdani, Maria Felicia Rega, and Mattia Sullini. 2025. *TRACE-ing the Gaps: Mapping Interventions on Incomplete 3D Meshes*. Zenodo. https://doi.org/10.5281/zenodo.15263277

[2] Balzani, Roberto, Sebastian Barzaghi, Gabriele Bitelli, Federica Bonifazi, Alice Bordignon, Luca Cipriani, Simona Colitti, et al. 2024. 'Saving Temporary Exhibitions in Virtual Environments: The Digital Renaissance of Ulisse Aldrovandi

[20] https://coara.eu/agreement/the-agreement-full-text/

– Acquisition and Digitisation of Cultural Heritage Objects'. Digital Applications in Archaeology and Cultural Heritage 32 (March):e00309. https://doi.org/10.1016/j.daach.2023.e00309

[3] Barker, Michelle, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, et al. 2022. 'Introducing the FAIR Principles for Research Software'. Scientific Data 9 (1): 622. https://doi.org/10.1038/s41597-022-01710-x

[4] Barzaghi, Sebastian, Alice Bordignon, Bianca Gualandi, and Silvio Peroni. 2024. 'Thinking Outside the Black Box: Insights from a Digital Exhibition in the Humanities'. In *Me.Te. Digitali. Mediterraneo in Rete Tra Testi e Contesti, Proceedings Del XIII Convegno Annuale AIUCD2024*, 138–42. Quaderni Di Umanistica Digitale. Dipartimento di Filologia Classica e Italianistica – Alma Mater Studiorum – Università di Bologna Associazione per l'Informatica Umanistica e le Culture Digitali – AIUCD. https://doi.org/10.6092/unibo/amsacta/7927

[5] Barzaghi, Sebastian, Alice Bordignon, Federica Collina, Francesca Fabbri, Bruno Fanini, Daniele Ferdani, Bianca Gualandi, Ivan Heibi, Nicola Mariniello, Arcangelo Massari, Marcello Massidda, Arianna Moretti, Silvio Peroni, Sofia Pescarin, Maria Felicia Rega, Giulia Renda, and Mattia Sullini. 2024a. 'A Reproducible Workflow for the Creation of Digital Twins in the Cultural Heritage Domain'. Zenodo. https://doi.org/10.5281/zenodo.14170838.

[6] Barzaghi, Sebastian, Alice Bordignon, Bianca Gualandi, Ivan Heibi, Arcangelo Massari, Arianna Moretti, Silvio Peroni, and Giulia Renda. 2024b. 'A Proposal for a FAIR Management of 3D Data in Cultural Heritage: The Aldrovandi Digital Twin Case'. Data Intelligence 6 (4): 1190–1221. https://doi.org/10.3724/2096-7004.di.2024.0061

[7] Barzaghi, Sebastian, Ivan Heibi, Arianna Moretti, and Silvio Peroni. 2024c. 'Developing Application Profiles for Enhancing Data and Workflows in Cultural Heritage Digitisation Processes'. In *The Semantic Web – ISWC 2024*, edited by Gianluca Demartini, Katja Hose, Maribel Acosta, Matteo Palmonari, Gong Cheng, Hala Skaf-Molli, Nicolas Ferranti, Daniel Hernández, and Aidan Hogan, 15233:197–217. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-77847-6_11

[8] Bordignon, Alice, Sebastian Barzaghi, Federica Collina, Francesca Fabbri, Bruno Fanini, Daniele Ferdani, Nicola Marinello, Arianna Moretti, Maria Felicia Rega, and Mattia Sullini. 2024. *Guidelines for the Digitisation of Museum and Art Collections*. Zenodo. https://doi.org/10.5281/zenodo.14249936

[9] Bordignon, Alice, Federica Collina, Francesca Fabbri, Daniele Ferdani, MARIA FELICIA REGA, and Mattia Sullini. 2025a. 'Balancing Photorealism and Transparency: A FAIR Approach to 3D Digitisation for Cultural Heritage Documentation and Dissemination', January. https://doi.org/10.5281/ZENODO.14639952

[10] Cyganiak, Richard, David Wood, and Markus Krötzsch. 2014. 'RDF 1.1 Concepts and Abstract Syntax'. W3C Recommendation. World Wide Web Consortium. https://www.w3.org/TR/rdf11-concepts/

[11] Demetrescu, Emanuel, Enzo d'Annibale, Daniele Ferdani, and Bruno Fanini. 2020. 'Digital Replica of Cultural Landscapes: An Experimental Reality-Based Workflow to Create Realistic, Interactive Open World Experiences'. Journal of Cultural Heritage 41 (January):125–41. https://doi.org/10.1016/j.culher.2019.07.018

[12] Derksen, Maarten, Stephanie Meirmans, Jonna Brenninkmeijer, Jeannette Pols, Annemarijn de Boer, Hans Van Eyghen, Surya Gayet, et al. 2024. 'Replication Studies in the Netherlands: Lessons Learned and Recommendations for Funders, Publishers and Editors, and Universities', January. https://doi.org/10.31219/osf.io/bj8xz

[13] Doerr, Martin, Christian-Emil Ore, and Stephen Stead. 2007. 'The CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing.' In , 51–56. https://doi.org/10.13140/2.1.1420.6400

[14] Doerr, Martin, and Maria Theodoridou. 'CRMdig: A Generic Digital Provenance Model for Scientific Observation', 2011. https://www.usenix.org/legacy/events/tapp11/tech/final_files/Doerr.pdf (accessed on 31/01/2025).

[15] Edmond, Jennifer, and Erzsébet Tóth-Czifra. 2018. 'Open Data for Humanists, A Pragmatic Guide', December. https://doi.org/10.5281/ZENODO.2657248

[16] Fanini, Bruno, Daniele Ferdani, Emanuel Demetrescu, Simone Berto, and Enzo d'Annibale. 2021. 'ATON: An Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage'. Applied Sciences 11 (22): 11062. https://doi.org/10.3390/app112211062

[17] Goodman, Steven N., Daniele Fanelli, and John P. A. Ioannidis. 2016. 'What Does Research Reproducibility Mean?' 8 (341).

[18] Gualandi, Bianca, and Silvio Peroni. 2024. 'Data Management Plan: Second Version', February. https://doi.org/10.5281/ZENODO.10727879

[19] Harpring, Patricia. 2010. 'Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA'. Art Documentation: Journal of the Art Libraries Society of North America 29 (1): 67–72. https://doi.org/10.1086/adx.29.1.27949541

[20] Karoune, Emma, and Esther Plomp. 2022. 'Removing Barriers to Reproducible Research in Archaeology', November. https://doi.org/10.5281/ZENODO.7320029

[21] Koster, Lukas, and Saskia Woutersen-Windhouwer. 2018. 'FAIR Principles for Library, Archive and Museum Collections: A Proposal for Standards for Reusable Collections'. The Code4Lib Journal, no. 40 (May). https://journal.code4lib.org/articles/13427

[22] Leonelli, Sabina. 2018. 'Rethinking Reproducibility as a Criterion for Research Quality'. In Research in the History of Economic Thought and Methodology,

edited by Luca Fiorito, Scott Scheall, and Carlos Eduardo Suprinyak, 36:129–46. Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009

[23] Moore, Jennifer, Adam Rountrey, and Hannah Scates Kettler. 2022. 3D Data Creation to Curation: Community Standards for 3D Data Preservation. ALA. https://www.alastore.ala.org/content/3d-data-creation-curation-community-standards-3d-data-preservation

[24] Niccolucci, Franco, Béatrice Markhoff, Maria Theodoridou, Achille Felicetti, and Sorin Hermon. 2023. 'The Heritage Digital Twin: A Bicycle Made for Two. The Integration of Digital Methodologies into Cultural Heritage Research'. Open Research Europe 3 (April):64. https://doi.org/10.12688/openreseurope.15496.1

[25] Peels, Rik, and Lex Bouter. 2018. 'The Possibility and Desirability of Replication in the Humanities'. Palgrave Communications 4 (1): 95. https://doi.org/10.1057/s41599-018-0149-x

[26] Penders, Holbrook, and De Rijcke. 2019. 'Rinse and Repeat: Understanding the Value of Replication across Different Ways of Knowing'. Publications 7 (3): 52. https://doi.org/10.3390/publications7030052

[27] Peroni, Silvio. 2017. 'A Simplified Agile Methodology for Ontology Development'. In OWL: Experiences and Directions – Reasoner Evaluation, edited by Mauro Dragoni, María Poveda-Villalón, and Ernesto Jimenez-Ruiz, 55–69. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-54627-8_5

[28] Rahal, Rima-Maria, Hanjo Hamann, Hilmar Brohmer, and Florian Pethig. 2022. 'Sharing the Recipe: Reproducibility and Replicability in Research Across Disciplines'. Research Ideas and Outcomes 8 (November):e89980. https://doi.org/10.3897/rio.8.e89980

[29] Riva, Pat, Maja Žumer, and Trond Aalberg. 2022. 'LRMoo, a High-Level Model in an Object-Oriented Framework', October. https://2022.ifla.org/

[30] Robinet, Fabrice, Remi Arnaud, Tony Parisi, and Patrick Cozzi. 2014. 'glTF: Designing an Open-Standard Runtime Asset Format'. In GPU Pro 5. A K Peters/CRC Press.

[31] White, Ethan, Elita Baldridge, Zachary Brym, Kenneth Locey, Daniel McGlinn, and Sarah Supp. 2013. 'Nine Simple Ways to Make It Easier to (Re)Use Your Data'. Ideas in Ecology and Evolution 6 (2). https://doi.org/10.4033/iee.2013.6b.6.f

[32] Wilkinson, Mark D., Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. Scientific Data 3 (March):1–9. https://doi.org/10.1038/sdata.2016.18

[33] Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. 'Good Enough Practices in Scientific Computing'. Edited by Francis Ouellette. PLOS Computational Biology 13 (6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510