ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Perspective

Cite this article: Hurdle K *et al*. 2025 A FAIR narrative on open science: when, who, what and how of computed tomography and the NoCTURN network. *R. Soc. Open Sci.* **12**: 251691.
https://doi.org/10.1098/rsos.251691

Received: 4 September 2025
Accepted: 7 November 2025

Subject Category:
Science, society and policy

Subject Areas:
e-science

Keywords:
metadata, findability, accessibility, interoperability, reuse, persistent identifiers

Author for correspondence:
Kelsi Hurdle
e-mails: khurdle@nyit.edu; khurds137@gmail.com

Supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8192660.

THE ROYAL SOCIETY PUBLISHING

A FAIR narrative on open science: when, who, what and how of computed tomography and the NoCTURN network

Kelsi Hurdle[1], Morgan Chase[2,4], Callie H. Crawford[5], Freya E. Goetz[6,8], Jaimi A. Gray[9], Alexander S. Hall[11,12], Jennifer J. Hill[7], John J. Jacisin III[10,13], Richard E. Johnston[14], Leigha M. Lynch[15], Elizabeth Martin-Silverstone[16], Heather F. Smith[15], Christopher M. Zobek[17] and Paul M. Gignac[18,3]

[1]Department of Anatomy, NYITCOM, Old Westbury, NY, USA
[2]Microscopy and Imaging Facility, and [3]Division of Paleontology, American Museum of Natural History, New York, NY, USA
[4]North Star Imaging, Rogers, MN, USA
[5]Department of Biology, Coastal Carolina University, Conway, SC, USA
[6]Department of Invertebrate Zoology, and [7]Scientific Imaging at the Micro Computed Tomography Imaging Center, Smithsonian Institution National Museum of Natural History, Washington, DC, USA
[8]Department of Marine and Environmental Biology, University of Southern California, Los Angeles, CA, USA
[9]Department of Earth and Planetary Sciences, and [10]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA
[11]Department of Materials and Structural Analysis, Thermo Fisher Scientific, Houston, TX, USA
[12]Dryad Digital Repository, Raleigh, NC, USA
[13]Department of Natural and Applied Sciences, University of Wisconsin-Green Bay, Green Bay, WI, USA
[14]Department of Advanced Imaging of Materials Core Facility, Swansea University, Swansea, Wales, UK
[15]Department of Anatomy, Midwestern University, Glendale, AZ, USA
[16]Palaeobiology Group - School of Earth Sciences, University of Bristol, Bristol, England, UK
[17]Department of Pathology and Anatomical Sciences, University of Missouri, Columbia, MO, USA
[18]The University of Arizona College of Medicine Tucson, Tucson, AZ, USA

KH, 0000-0003-3799-1349; MC, 0000-0001-7486-7174; CHC, 0000-0002-7225-8137; FEG, 0000-0001-5696-4333; JAG, 0000-0002-4758-8572; ASH, 0000-0002-3807-1395; JJH, 0000-0002-5550-8173; JJJIII, 0000-0002-9422-9142;

REJ, 0000-0003-1977-6418; LML, 0000-0003-3055-8438; EM-S, 0000-0003-0139-2109; HFS, 0000-0002-3738-0903; CMZ, 0000-0002-1482-0726; PMG, 0000-0001-9181-3258

Open science, the practice of making research products and processes free and available to scientists and non-scientists alike, is transforming image-based studies using non-clinical computed tomography (CT). Recently, the National Science Foundation has established research coordination networks to help address principles of findability, accessibility, interoperability and reusability (FAIR) within the United States (US) scientific enterprise and in partnership with the global community. The Non-Clinical Tomography Users Research Network (NoCTURN) is one such organization, focused on CT data accessibility and reuse, helping to address significant obstacles related to bespoke hardware, proprietary software and siloed workflows. NoCTURN strives to develop and maintain linkages across CT ecosystems capable of driving cultural and technological changes. To that end, we discuss the last two decades of FAIR principles (*when*) and present our recommendations for CT data access by focusing on a survey of predominantly US CT users (*who*), the definition and role of metadata schemas (*what*), and the broad use of persistent identifiers (*how*).

1. Introduction—the 'when' of open science

Recently, global scientific communities have been engaged in conversations on establishing broad, transparent science standards, the overarching goal of which is to facilitate wide dissemination of scientific output, both in the form of published papers and underlying data. Organizations such as the European Commission [1], G20 [2] and the United Nations Educational Scientific and Cultural Organization [3] have endorsed inclusive transparency in scientific practices, setting the stage for pro-open science policy changes. Collectively, these organizations promote the public benefits that result from global scientific cooperation as open science: 'the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity' [4]. Open science is, thus, a philosophy that facilitates and promotes the transparency of scientific knowledge creation and dissemination to ensure such knowledge is findable, accessible, interoperable and reusable (FAIR) within and outside the scientific community [5]. Findability refers to ensuring that data and metadata are easily discoverable. For example, a researcher doing a study on the shapes of raccoon teeth should be able to locate computed tomography (CT) scans for other raccoon specimens relatively easily. Accessibility refers to minimizing hurdles to acquiring data, such as ensuring the user has access to any necessary passwords or login information. The first step in this process is the data being uploaded to a repository (e.g. a digital database for information storage, which can be public or private) where it can be accessed by others. In addition, even if a password or permissions are required for download, this information should still be available for potential users (see §3.2). Interoperability refers to the integration of processes and knowledge with other data, programmes and workflows. In other words, the data must be in a format that is widely usable and/or convertible. For example, most users store data in standardized file types such as .tiff or .dcm, rather than proprietary software which may require an expensive purchase. Reusability refers to optimizing how well data and knowledge can be repurposed, which requires that the data and metadata be thoroughly described for replicability. For example, a researcher wishing to incorporate available data into their study should be able to select data which they know has similar scan parameters to their planned study.

The implementation of open science principles is particularly relevant to the democratization of fields that generate large, shareable datasets. A leading example is CT, which has revolutionized scientific investigation by revealing the three-dimensional (3D) internal structure(s) of objects at unprecedented spatial resolution and volumetric fidelity. Tomographic imaging produces virtual stacks of image data (slices) and resulting 3D models. These vast amounts of high-resolution scan data can exceed hundreds of gigabytes or even terabytes. This has also brought challenges relating to the management and dissemination of datasets that can be unobtainable or practically unshareable. In addition, scanner outputs may contain fully or partly proprietary file types, which can silo datasets into vendor-specific software pipelines. These sharing and accessibility challenges must be overcome to ensure the transparency of CT data through FAIR practices.

In an effort to meet these challenges, several initiatives have led to the creation of large-scale, openly accessible online databases that work directly and indirectly with data aggregators and repositories [6–13]. These repositories typically house digital representations of physically difficult-to-access, priceless, one-of-a-kind specimens from natural history, medical, cultural, anthropological and fossil collections from around the world. A major component of repository data management is linking physical objects with these digital representations and tracking their access within repositories via persistent identifiers. As 3D capture technology becomes more advanced, an increased demand for high-resolution virtual specimen data will follow. This trend will necessarily drive data image curation and the use of persistent identifiers for data and metadata [14], which will be critical for customizing and standardizing CT imaging workflows to ensure scan data are tested against, verified for and compatible with multi-modal archive resources [15].

Over the last two decades, the drive to realize these best practices has gained momentum globally, with governmental initiatives and policies aiming to ensure wider access to publicly funded research. Here we summarize how these policies have evolved over time and borrowed from each other to address data accessibility challenges and promote the FAIR principles to increase equity in science.

1.1. European Commission open data policies

The movement towards open science gained momentum in 2011 in Europe with steps taken by the European Commission to publicize its documents within a European Union (EU) Open Data Portal [16]. This set the stage for a larger community shift, and in 2012 the Commission published a formal recommendation 'on access to and preservation of scientific information' [17]. The recommendation outlined open access expectations for scientific publications and research data resulting from publicly funded research. From 2014 onwards, the European Research Council's (ERC) Horizon2020 initiative took additional steps to advance open science in Europe through an Open Research Data Pilot programme. It required all ERC-funded publications to be publicly accessible via open access—notably with an opt-out provision. This provided an attempt to balance openness with commercialization, privacy and other concerns [18,19]. Following the data pilot programme, the transition to Horizon Europe in 2021 solidified these commitments. Under a newly minted principle—'as open as possible, as closed as necessary'—all EU member states have been invited and encouraged to participate in the publicly facing Data.europa.eu portal [18,19].

1.2. The Concordat on Open Research Data and United Kingdom Research and Innovation (UKRI) Open Research Policy

Confluent with Horizon 2020 in Europe, a coalition of funders and higher education organizations in the United Kingdom (UK) published the Concordat on Open Research Data (ORD), establishing guiding principles and best practices [20]. However, the 2016 Concordat did not mandate policies and was perceived to be unaligned with FAIR principles that otherwise champion recognition, incentives and clear expectations. To address these concerns an ORD Task Force was established, culminating in the development of more robust ORD policies for the UK (ORD Task Force, 2018 [21]).

Following the passage of the 2016 UK EU membership referendum, UK organizations began managing advances in FAIR and open science in parallel to the EU. For example, in 2021 the UK Research and Innovation (UKRI) organization first published their Open Access Policy, based on recommendations from the ORD Task Force. This policy was updated in 2023 to require all published research that acknowledges funding from the UKRI, or any of its constituent funding bodies, to be published under Creative Commons BY terms. These terms allow the public to use the work with proper attribution and adhere to FAIR Principles. Notably, this has been supported by block grants via UKRI to enable open access publishing, which has resulted in a free and open tier at a significant number of academic journals.

1.3. Findable Accessible Interoperable Reusable Open Science (FAIROS) initiative and the Nelson Memorandum

Contemporaneous with the EU Open Data Portal, the United States (US) Office of Science and Technology Policy (OSTP) published a memorandum in 2013, which began the process of expanding sharing requirements for publicly funded research [22]. The memo, entitled 'Increasing Access to the Results of Federally Funded Scientific Research', laid the groundwork for current federal data-sharing policies but had limitations in scope, applying only to agencies with significant research and development budgets and lacking specific requirements regarding accessibility and timely public release. These policies were succeeded in 2022 by the OSTP Nelson Memorandum ('Ensuring Free, Immediate and Equitable Access to Federally Funded Research' [23]). This update included two core advancements in open science policy for federally funded peer-reviewed research: (i) elimination of the previous 12 month embargo for federally funded research, meaning that affected research articles must be open access immediately upon publication; (ii) requirements for underlying data from federally funded projects to also be made immediately available.

In response to the Nelson Memorandum, the US National Science Foundation (NSF) released a programme solicitation for grant applications to establish Findable Accessible Interoperable Reusable Open Science Research Coordination Networks (FAIROS RCNs; NSF 22-553). The solicitation aligns with NSF's mission to facilitate research that benefits the public. The primary goal of the FAIROS RCN programme was to develop and support research networks of scientific communities that focused on advancing FAIR and open science principles within and between fields. By supporting research networks to address hurdles and develop plans for catalysing open access in their disciplines, FAIROS RCNs help to reduce barriers for accessing scientific knowledge and thereby increase equity of access to all.

2. The 'who': NoCTURN's community

One such FAIROS RCN is the Non-Clinical Tomography Users Research Network (NoCTURN; https://nocturnetwork.org/), which seeks to advance FAIR and open science practices in the domain of CT imaging. NoCTURN was established as a scientifically cross-cutting organization bringing together experts from across research, industry, education and data repositories to address the fragmented and opaque nature of knowledge dissemination within the CT community. The network has successfully generated resources and guidelines for ensuring FAIR CT data [10]. This objective has been accomplished through a grass-roots effort to establish what CT users aim to make FAIR, how they envision accomplishing this goal, and who will be the greatest beneficiaries of CT-generated information.

NoCTURN members are spread across North America, South America, Europe and Africa. Affiliation information is provided once someone new joins the network but no additional facts about individuals are given at that time. To better appreciate the backgrounds of its members, the network conducted a survey in 2023 of its North American members (IRB protocol #ET00022672). This was a first step towards characterizing the community (figure 1), with an international survey planned for the future. Based on the results (response rate of 33%), NoCTURN members average 11 years of experience with non-clinical tomography. The majority are employed or study at universities, colleges and museums. The remainder includes CT professionals employed in industry, government or national labs and medical centres. Our collective expertise encompasses a wide range of career experience: professors (47%), lab or research technicians (15%), graduate students (13%), lab managers or directors (8%), postdoctoral and visiting researchers (8%) and commercial roles (4%). The survey indicated that 79% of responders work at a college or university, 17% work in a museum, 8% in government or national labs, 6% in industry and 6% indicated 'other' (note, categories are non-exclusive). As their highest level of education, most (70%) responders say they have a doctoral degree, 15% have a master's degree and 15% indicate having a bachelor's degree. Demographically, despite the level of education and high percentage of tenure-track professionals, the survey respondents skew younger, indicating early- and mid-career participation: 18–26 (9%), 27–35 (28%), 36–44 (38%), 45–53 (15%), 54–62 (8%) and 63+ (2%). Though imperfect in its opportunistic collection, this survey paints a picture of a community of world-leading experts in producing, using, and managing CT data, alongside training others to produce and work with it as well. Nonetheless, this effort is important for enumerating the need to bring new demographics and additional perspectives into the community. Indeed, this is a major objective for the next phase of NoCTURN network growth.

Given the breadth of career paths taken by members of the community, an important role for NoCTURN is the networking and mentoring of its members. For example, a community of public and private innovation has flourished around CT technologies, as demonstrated by collaborative advancements in X-ray CT scanners through academic research; innovative neutron CT techniques from joint government–industry projects; and cutting-edge synchrotron-based imaging breakthroughs via national lab-private sector collaborations. This affords the opportunity for CT specialists to move between academic and industry professions, which are often considered niche in other fields. Few resources exist to guide and support these transitions, which are further complicated by the scores of machines, output formats, file types, data processing pipelines and data management options that are used daily by CT labs around the world. Because labs undertake these actions in parallel, the community benefits that would normally come from exchanging knowledge has been limited without a professional network such as NoCTURN. Historically, operators have informally handed down knowledge and protocols that have been adapted as equipment and software are updated and new techniques are integrated into operating procedures. To address the disconnect, NoCTURN draws from these experiences to better illuminate career insights and facilitate its members' career development. For instance, NoCTURN's Open Science Committee includes scanner operators who identified the need for web-based CT-facility help resources aimed at newly established facilities and personnel. The webpage includes FAQs alongside lists of relevant information for establishing and maintaining a CT lab. It is now published on NoCTURN's website for public access (https://nocturnetwork.org/resources/ct-help-page/).

Whereas NoCTURN focuses on the frameworks of CT data science and facilitating communication, our members are also scientists who use CT technologies to visualize, quantify and analyse physical objects. To this extent, NoCTURN—a research coordination network—is unable to solely fully address all the needs of our members. Within microscopy sciences broadly, organizations such as the Royal Microscopy Society (https://www.rms.org.uk/), Bioimaging North America (https://www.bioimagingnorthamerica.org/) and the International Society for Neutron Imaging (https://www.isnr.de/)—research consortia—have come together explicitly to advance the boundaries of human knowledge. For the past decade, the non-clinical CT community has organized the specialist conference Tomography for Scientific Advancement (ToScA; https://www.toscainternational.org/) affiliated with the Royal Microscopy Society to achieve this end. ToScA conferences have been valuable ways to cross-pollinate approaches to CT research, training and outreach across disciplines and industries. Starting in 2025, NoCTURN has joined ToScA as a partner organization. Together, they serve a common community with distinct aims of cultivating the expertise and background to improve the coordination of CT technologies, innovations and STEM education, while simultaneously using CT science to address hypotheses about the organization of the natural world.

3. Metadata in computed tomography: the 'what' of open science

Metadata refers to data that provide information about other data [24–26] by helping to describe, explain or give context to the primary data. Generally, this makes primary data easier to organize, find and interpret. For example, metadata associated with research articles includes key information that summarizes the study (e.g. the title, authors, publication date, abstract, keywords and journal name). Within the realm of clinical CT imaging, metadata serves a larger role in capturing important patient-relevant information alongside scanner parameters and technical schema. Efforts by the Digital Imaging and Communications in Medicine standards committee and its working groups have established standardized metadata categories that incorporate discipline-specific roadmaps (e.g. for cardiac medicine, nuclear medicine, radiotherapy; see [27] and references therein) that meet the bespoke needs of imaging patients under a wide range of clinical conditions. These needs differ from those of metadata associated with non-clinical CT imaging, which is often limited to specimen identification information and scan parameters only. For example, non-clinical CT stack metadata includes information about the sample, image capture and reconstruction steps such as specimen details (e.g. species moniker, sample catalogue number, experimental parameters, preparation history), scan set-up information (e.g. number of images captured, frame averages, filter details, exposure timing), hardware specifications (e.g. instrument make and model, persistent identifier, detector size, source type) and additional software details (e.g. application name, version number, post-processing

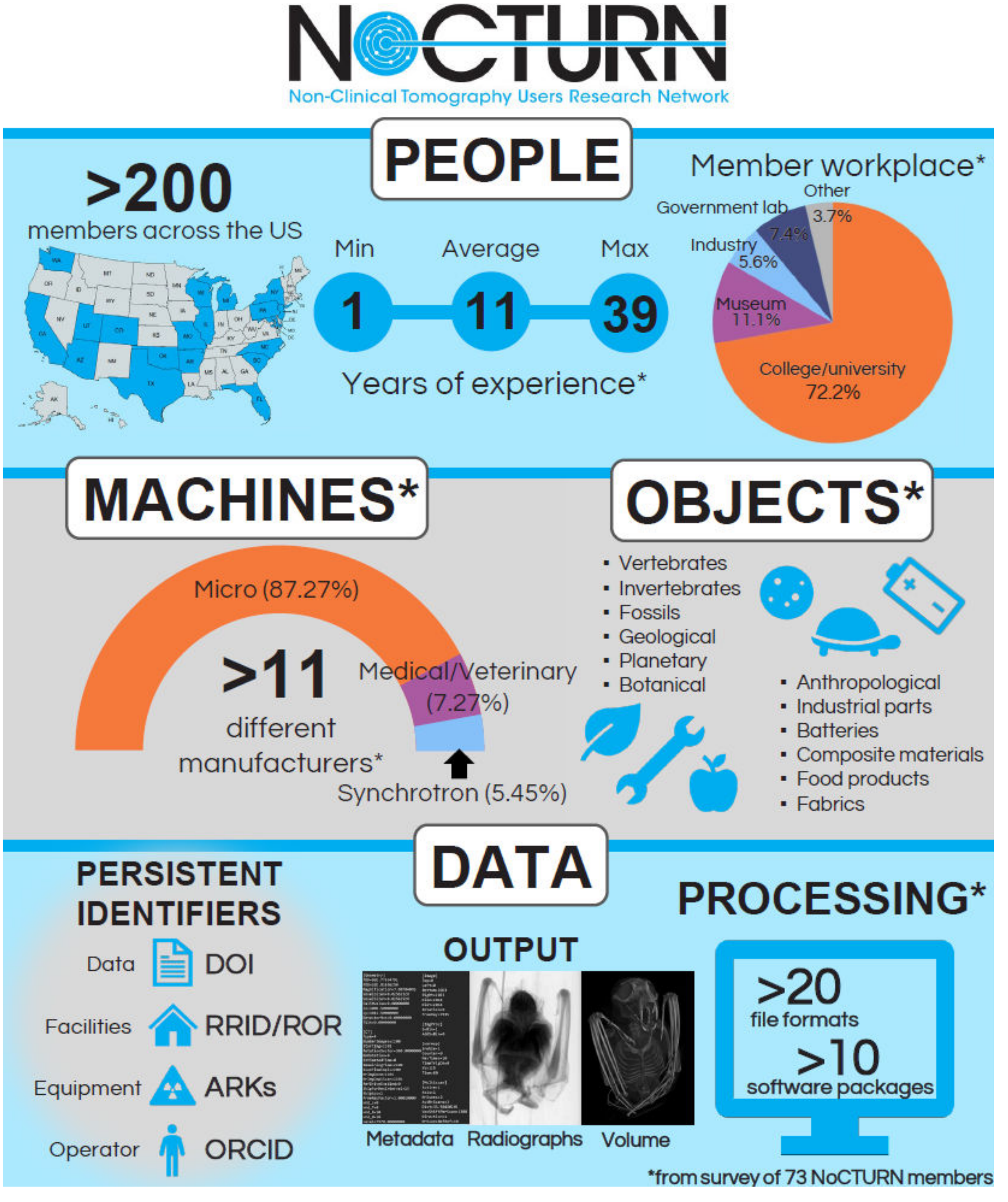

Figure 1. NoCTURN network graphic displaying the essential components of the field of non-clinical CT. All information was directly taken from the 2023 survey.

filters). Notably, some of this information is actively collected by the scan operator at the time of scan, but much is recorded automatically by the scanner software and stored in one or more Unicode-8 text files (UTF-8).

Appropriately documented metadata explains 'what' was done to the sample to capture the CT visualizations [28]. Metadata are crucial for the broad community of specialists interested in using CT imaging [29] because they provide context for descriptive and quantified analyses, digital object reconstruction and scientific discovery. CT operators refer to metadata to replicate scans [5,27,30]. Researchers and their trainees utilize the information from metadata to ensure accurate analysis and 3D renderings of scanned objects [31]. Manufacturers test the operation of CT equipment and read the results through metadata [32]. Developers improve and test their products with the use of metadata, optimizing software and computational efficiency with reference to the primary data as well as the metadata contents [33].

Replicability is the cornerstone of scientific advancement. Although CT-based metadata provide actionable information, they are not organized following a replicable standard. For example, each manufacturer may record different information within their metadata. This includes unique text fields and bespoke ways of representing the same kind of information, which limits the use and reuse of CT scans by erecting barriers to interoperability. This represents a replicability issue, because well-organized metadata are otherwise required for CT operators to recreate datasets. To better facilitate this need, Wilkinson *et al.* [5] argues that metadata should support manual and automated operations that promote FAIR standards. In alignment with these goals, we recommend that community action is needed to address metadata best practices in non-clinical CT by developing semantic metadata standards (e.g. structured to be human and machine readable).

3.1. Reconstruction and 3D rendering

As with other technical fields, understanding file formats and metadata related to CT imaging is pivotal for effective data management. Here we catalogue file formats, extensions and uses in CT image capture and processing. For example, the choice of image format significantly influences how data are processed, stored and retrieved. Image slices may be exported in a number of file formats such as .tif, .bmp, .jpeg, .dcm or .png. Formats like .tif and .dcm are especially common—the former is excellent for high-resolution imaging owing to its lossless nature, while the latter additionally offers structured metadata embedded within files, providing detailed context for datasets. Typically, a manufacturer's reconstruction programme requires one or more complete metadata files for successful reconstruction. Once finished, the resulting image slices compose a volume that can then undergo data analysis.

Certain reconstruction programmes include analytical features, but that is not their primary function. Therefore, the image slices are typically loaded into additional programmes for 3D rendering. These platforms are specifically designed for multi-planar viewing, creating 3D virtual objects and advanced data analysis (see table 1 for commonly used examples). There are numerous 3D rendering and analysis software platforms available for tomographic data, including fully proprietary, free with limited functionality, free and fully functional for non-commercial use, and open-source. Proprietary software—programmes provided by the same manufacturer of the CT instrument—are specifically coded to utilize metadata files written by the same company. Some manufacturers provide access to these tools on a limited basis or to non-commercial users (see table 1). By contrast, free and open-source software (FOSS) makes the programme code openly available to the public, but may not be interoperable with metadata in non-standard or proprietary formats.

3.2. Common repositories and metadata storage

A major benefit of electronic resources has to do with their dynamic nature; they are constantly evolving as technology expands opportunities for adaptability and reuse. Tracking these changes over time ensures that users can update records consistently while preserving information about past parameter choices and actions. When issues arise—whether during software testing or user interactions with specialized tools like CT segmentation—metadata serves as a reference point for troubleshooting, making these files especially important elements in the context of data repositories. According to the Registry of Research Data Repositories (re3data.org), over the last decade, the number of available public access repositories has increased from 1821 in 2015 to 3251 in 2024. Thus, repository popularity is allowing data publishers to easily share their data and metadata with the public in accordance with the Nelson Memorandum [23] in the US and similar aforementioned initiatives globally.

Data repositories are common database tools for archiving data and metadata as well as extremely popular methods of sharing imaging data among the non-clinical CT community. There are many data repositories to store data such as clinical and non-clinical CT scan data, figures, supplementary data files, big data, geospatial data and code. Commonly used repositories include generalist databases for broad accessibility, journal publishers for academic contributions, curated collections for interrelated datasets, field-specific repositories for specialized data and digital museums. A list of those most commonly used by the non-clinical CT community can be found in table 2.

Of these repositories, members of the NoCTURN community have tended to favour MorphoSource, specifically, for its metadata-focused manifests, specimen-oriented approach to curation and

Table 1. Examples of proprietary and free and open-source software (FOSS) for graphical user interface (GUI)-based CT reconstruction, 3D rendering and analysis tools commonly used by NoCTURN members. (*Indicates analytical capabilities as well as 3D rendering toolset.) A list of recommended software used by synchrotron facilities can be found at https://tomopedia.github.io/software/.

programme	proprietary or FOSS	3D rendering*	analytical software
Adams	proprietary	yes*	no
Avizo/Amira	proprietary	yes*	yes
Drishti	FOSS	yes	no
Hyperworks/Hypermesh	proprietary	yes*/no*	no
Imalytics Preclinical	proprietary	yes*	yes
Inspect-X	proprietary	yes	no
Mimics	proprietary	yes*	yes
NRecon	proprietary	no	yes
Volume Graphics Studio Max	proprietary	yes*	yes
Dragonfly	proprietary, free for non-commercial use	yes*	yes
OsiriX	proprietary, limited functionality free version	yes	yes
3D Slicer	FOSS	yes*	yes
ImageJ/Fiji	FOSS	no*	yes
SPIERS	FOSS	yes*	yes
Syglass	proprietary	yes	yes

comprehensive response to user feedback. Metadata is accessible when it is free to obtain. It should also be permanently available and accessible without barriers. MorphoSource, for example, uses an open access web interface without proprietary plugins or code. Notably, to upload data and metadata, the user must sign up for an account with a unique login before inputting acquisition and reconstruction information. This protects against data scrapers and bots on the platform; however, it is notable that required credentialing does add friction to data accessibility for both data generators and reusers, but it is a common requirement of data repositories. The metadata information that must be appended to each uploaded two-dimensional (2D) or 3D dataset minimally include: image acquisition details, identifiers, permissions, object details, collection information and other general details [7,14]. Importantly, for CT data stacks to be uploaded to the database and downloaded by users for reuse, the data owner must also select a data manager, download reviewer and license agreement. Once selected for download, the dataset is output as one or more .zip files, which includes metadata .csv and .xlsx spreadsheets [14]. The spreadsheets provide users an organized schema of data and metadata information that was uploaded by the owner of the media files from the original metadata output.

3.3. Metadata files are easily transportable and can be opened as text files

Metadata files produced after scanning are typically small or low volume (i.e. less than or equal to 1 megabyte) and easily shareable, allowing the best practice of uploading all metadata information whenever repositing media. An important caveat is that any metadata not uploaded to the repository cannot be easily accessed by reusers. Metadata are often saved as file types designated for specific scanning software (i.e. .pca, .xkekct, .log, etc.); however, most, if not all such file types, can also be opened as UTF-8. This feature makes it relatively easy to check the settings used in the instrument and during initial reconstruction steps. The small file size makes the metadata files more easily transmittable between institutions without complications. Universities, museums and vendors utilize data storage and transfer sites such as Dropbox, Google Drive and Microsoft OneDrive to quickly share their metadata information with each other. Thousands of metadata files are small enough to be stored long-term on a low-capacity storage website, an external drive or a local computer hard drive. By contrast to the metadata files, projection images and the volumetric or tomographic slices created from them are substantially larger. On a daily basis hundreds of gigabytes of scan data are produced by laboratory scanners and terabytes of data from synchrotron facilities (see [34] for facility

Table 2. Exemplar information repositories that allow CT or CT-derived data identified by name, category and URL. Visit the URLs or see the NoCTURN data-repository website (https://nocturnetwork.org/resources/data-repositories/) for further documentation and repository comparisons. *Indicates active or planned repository for CT data and metadata.

repository name	type of repository	URL
Dryad	generalist	https://datadryad.org
Figshare	generalist	http://figshare.com
Zenodo	generalist	https://zenodo.org
Astromaterials 3D	curated collection	https://ares.jsc.nasa.gov/astromaterials3d/
Aves3D	curated collection	https://www.aves3d.org
Cipher Machines	curated collection	https://ctview.dmd.zone/
NASA PSI Public Data	government associated	https://doi.psi.ch
Smithsonian 3D Digitization (e.g. Voyager, Cook)*	institution associated	https://3d.si.edu
Harvard's Dataverse	institution associated	https://dataverse.org
University of Michigan Deep Blue Repositories*	institution associated	https://deepblue.lib.umich.edu/data
DigiMorph*	organization associated	http://www.digimorph.org/index.phtml
IEEE DataPort	journal publisher	https://ieee-dataport.org
MorphoMuseum	journal publisher	https://morphomuseum.com
Composites Library	topic-specific	https://ct-composites.nottingham.ac.uk/
FaceBase*	topic specific	https://www.facebase.org
GigaDB	topic specific	https://gigadb.org/
LifewatchGreece micro-CT*	topic specific	https://microct.portal.lifewatchgreece.eu/
MorphoDBase	topic specific	https://www.morphdbase.de/
MorphoSource*	topic specific	https://www.morphosource.org
Palaeontology Database	topic specific	https://paleo.esrf.eu
Phenome10K	topic specific	https://www.phenome10k.org
Tomo-Meta	topic specific	https://tomo-meta.readthedocs.io/

categories and data volume breakdown). Managing and sharing such large volumes of scanned data can be prohibitive. Thus, rather than transporting an entire volume of CT data, a user can instead transfer and store a metadata file and use its information to replicate the set-up for a prior scan with maximum fidelity.

3.4. Differing metadata conventions limit interoperability and hinder reuse

Differing metadata conventions exist within non-clinical tomography because there is a 'lack of consensus and clarity in how fields are used or defined' by data repositories, data publishers, CT operators and manufacturers [35]. Metadata files are routinely produced as a result of the scanning process and contain essential information for replicating a scan. However, there is inconsistency in the types of information that are recorded and reported. While researchers generally maintain logs of the steps taken with raw and reconstructed data, the fine spatial measurements of the scan volume and the settings applied by the machine operator may not necessarily be included as an output alongside the media [34,36,37]. Manufacturers commonly design their CT equipment to generate a metadata file using their own terminology and file type. For instance, Bruker Skyscan systems automatically generate a .log file after every scan that contains detailed information about what parameters were selected during set up such as the voltage, current, exposure, rotation and steps. Custom metadata information cannot be entered by the operator. By contrast, Waygate Technologies (formerly GE/Baker Hughes) Phoenix v|tome|x system generates five metadata files, each containing a portion of information pertaining to the identification of the specimen or sample, scan parameters and reconstruction. The operator can also manually enter specimen metadata which become part of the metadata files generated during the scan. As a result of this discordance, information reported by multiple individuals or organizations across different scanner platforms, opportunities for error and miscommunication are high [38]. Altogether, this lack of metadata standardization can make maintaining information consistency needed for achieving interoperability difficult.

This inconsistency is exacerbated by both incongruous terminology similarity and extensive parameter-name differences that exist between hardware platforms for the same parameters. Dozens of CT scanner hardware configurations are utilized across the NoCTURN network, and this ambiguity has led to difficulty for novices and experts alike when discussing scanning standardization. For example, closely related words are used to describe exposure timing and number of exposures across different systems. Specifically, the number of backscatter images captured during a scan is sometimes called 'number of exposures', 'number of projections' or 'rotation step'. Whereas, the length of the exposure in milliseconds is sometimes referred to as 'exposure', 'timing' or 'fps' (see table 3). Importantly, if the value of exposure timing is incorrectly entered into the exposure number field during the 3D reconstruction process [39], the discrepancy causes the creation of an image volume to fail. In a similar, common data entry issue, data analysis software requires voxel dimensions. If inaccurate spatial dimensions are used, then the resulting output lacks spatial accuracy and has little quantitative value. Even in cases when most metadata are in hand, replicating the original image capture is a challenge, or not possible, without the full inclusion of parameters chosen by the original scan operator during the initial data acquisition. Thus, there is a high risk for communication errors among users especially during problem-solving situations when precise meaning is critical and most challenged by imprecise and non-standardized vocabulary. To address this issue, we recommend a standardization of terminology based on a commonly used set of terms and phrases within metadata. In an attempt to remove the ambiguity surrounding CT terminology, members of NoCTURN's Interoperability Working Group deliberated on the use of common metadata terms and came to an agreement on accepted definitions, which have been compiled as the NoCTURN website terminology list (https://nocturnetwork.org/terminology/). The list also provides additional information such as references that were used to support the use of vocabulary standardization.

3.5. Proposed solution

Herein we have identified gaps relating to the use of common terminology, non-uniform conventions for organizing metadata contents, and inability for 3D rendering software platforms to access metadata information. To resolve these problems, we recommend that manufacturers enable users to produce a single, universal metadata file. To accommodate the wide range of samples captured with CT imaging and to acknowledge potential future additions to CT metadata contents, we recommend a two-tiered approach to codifying the universal metadata file. This universal metadata file should have a well-structured and defined interface for entering metadata, wherein primary fields would be identical and mandatory (e.g. a drop-down selection list displaying standardized vocabulary [38]). If a piece of information is necessary to reproduce or interpret the scan, then it needs to be included. For example, Stock & De Carlo [34] and Davies *et al*. [40] provide a minimum standard for this purpose that aligns with the Nelson Memorandum [23]. Secondary fields, such as those not needed to reproduce the scan itself but representing discipline-specific conventions (e.g. geographic information system coordinates, project titles, persistent identifiers), should be customizable and optional. Secondary fields would not necessarily be interpretable by all software accessing the metadata, whereas primary fields would be. To help transition our recommendation into practice, we suggest allowing manufacturers and users to retain their current metadata conventions, while also having a universal metadata file that can be machine and human read for broad transfer between systems and users. Over time, we anticipate that the usefulness of a universal metadata file will draw ubiquitous adoption. If standardization of metadata can be achieved this way, it will facilitate unprecedented interoperability and reuse, thus supporting the significance, growth and future directions in CT research with a negligible addition to data storage requirements.

4. Persistent identifiers (PIDs): the 'how' of open science

Metadata are findable through the use of persistent identifiers (PIDs) [41], which are a series of unique letters and numbers that serve as a permanent reference to a digital resource (e.g. journal article, dataset, volume file, web page, CT scan, image, coded script, etc.). These often follow an encoding

Table 3. Examples of non-common terms used by CT manufacturers and data repositories for the same parameters. An explanation of each term can be found on the NoCTURN website terminology list.

NoCTURN website	MorphoSource	Zeiss	GE/Waygate	Bruker	Nikon	North Star
frame averaging	frame averaging	images per projection	average (Avg in .pca)	averaging (frames)	averaging	frames averaged
frames/exposures/projections	number of projections	number of Projections	images (NumberImages in .pca)	rotation step (deg)	projections	number of projections
kVP/voltage	voltage	voltage	voltage	voltage	XraykV	voltage
timing/exposure time	exposure time	exposure	timing	exposure	exposure	Fps

system, but can be assigned arbitrarily. A PID connects a name with a digital or physical information resource such that active maintenance of a PID within a transparent online database ensures that the information resource is traceable. Indeed, PIDs are required to facilitate effective interaction with archived digital resources at scale, and, thus, the included information needs to be accurate (e.g. verified against a CURATED (Check, Understand, Request, Augment, Transform, Evaluate, Document) rubric [42]), transparent and accessible. These characteristics help to prevent misinterpretation, uphold data FAIR-ness activities, ensure timeliness (see real-time data cost benefit analysis [43]) and make the information easy to understand (e.g. machine-readable) [44,45]. Altogether, these aspects of PIDs enable data—and research—reproducibility [46].

As PIDs continue to be an integral part of the research process, several countries (e.g. Canada, Australia, UK, Europe and South Korea [47–51]) have adopted regional and national PID strategies. These initiatives aim to create inclusive communities that foster openly transparent and consistent use, infrastructural support and deployment plans. As we outline below, these strategies include precise details on use-cases to clarify who should use PIDs, why PIDs are critically important to data implementation, when it is appropriate to use such identifiers and in what context PIDs are likely to be most effective [52–55].

Important PIDs within the CT community include Research Resource Identifiers (RRIDs), Research Organization Registries (RORs), Archival Resource Keys (ARKs) and Open Researcher and Contributor IDs (ORCIDs). Many of these identifiers link critical information that addresses issues in metadata standardization, improve interoperability and make CT data more discoverable and usable [56–58]. As table 4 summarizes, RRIDs are increasingly used to reference integrative research resources and imaging facilities, thus ensuring traceability of tools and resources, although they require facilities to be actively maintained and possess a public-facing website. For example, NoCTURN publishes gold standards and best practices for CT data accessibility, and it has an RRID (RRID:SCR_025089) for referencing all affiliated in-network-created research resources.

RORs are issued to research organizations. These PIDs verify and trace entities through institutional changes like mergers or rebrands, and integrate with metadata systems such as ORCID and Crossref to map user affiliations (e.g. FREYA project). ARKs are used to reference digital, physical or conceptual information objects, and are often applied as equipment IDs or lab configurations. ARKs support long-term access, are low-cost, globally resolvable and provide versatile ways to track the 'what' of research activities. Finally, ORCIDs are unique 16-digit identifiers for researchers, linking their work, publications and grants to ensure clear attribution across their career. ORCIDs are used globally across disciplines and can be integrated with other PID systems to create interoperable records.

In an attempt to make PIDs more equitably accessible and discoverable to various communities, digital object identifier registration agencies (DOI-RAs) like CrossRef and DataCite, have partnered with ORCID to develop a US National PID strategy. The creation of a centralized PID infrastructure would underpin a unified approach to research management that addresses diversity and fosters cooperation through community participation. This approach would enhance research assessment, provide transparency and accountability, and encourage global collaboration, while working to alleviate administrative burden, specifically the time and costs that are often associated with data management [50]. Although this strategy outlines the overall benefits of PIDs and their associated

Table 4. Notable persistent identifiers (PIDs), what they identify, their benefits, limitations and stakeholders within NoCTURN.

persistent identifier (PID) type	categories identified	benefits	limitations	stakeholders
Research Resource Identifier (RRID) https://www.rrids.org/current-project	biological research resources, CT facilities, scanning equipment, reconstruction and visualization tools	ensures traceability, creates accountability, easy to apply and free	requires public-facing website, facility must be actively maintained, approval process can be delayed	researchers, lab managers, funders, software developers, technicians
Research Organization Registries (ROR) https://ror.org/	research organizations, affiliations, infrastructure projects	verifies organizations, supports metadata tracking, open API access	less suitable for legacy organizations, integration issues	publishers, funders, libraries, data managers, repository representatives, policy administrators
Archival Resource Keys (ARK) https://arks.org/about/	digital, physical or conceptual information objects, lab/facility configurations	low-cost, globally resolvable, versatile, can represent future objects	requires maintenance, may be complex for small organizations	researchers, archivists, data managers, lab managers, repository representatives
Open Researcher and Contributor IDs (ORCID) https://orcid.org/	researchers, contributors, work publications, grants, education and employment records	connects researchers to their work, interoperable, globally recognized	requires active management, may overlap with other PID systems	researchers, publishers, funders, repository representatives, policy administrators

metadata, and provides a thorough framework for PID stakeholders, we note that it is centred around digital object identifiers (i.e. DOI-centric, excluding information resources beyond publications or literary contributions; see fig. 1 of [59]). Examples include software, source code and the extensive work undertaken by the Research Data Alliance/Force11 Software Source Code Identification Working Group [60], FAIR4RS (FAIR for Research Software) Working Group [61], Software Heritage Project (softwareheritage.org) and SWHIDs (Software Hash persistent IDentifiers [62,63]). The US National PID strategy may therefore ultimately require amendment to include identifiers relevant to the data collected with advanced imaging tools and other cutting-edge research hardware and software.

The use of PIDs in CT ensures the longevity and open accessibility of data and metadata. PIDs operate to link researchers via an ORCID to (i) their scientific contributions (e.g. publications, protocols, patents), (ii) their relevant research affiliations (e.g. commercial, industrial, public and private facilities, laboratories, institutions and libraries) often through the use of RRIDs or RORs, and (iii) the technology, equipment and visualization tools (e.g. μCT and MRI scanner(s), data network server(s), open source and commercial licensing software) that collected or analysed the physical specimen, inclusive of its locality, inherent characteristic attributes, intrinsic properties and provenance via ARKs. It is both necessary and imperative that unique persistent identifiers are applied to both physical-tangible objects and the digital object representations or iterations (e.g. 3D CT scan, 2D tomographs, photogrammetry images, meshes, point clouds, rendered models, etc.) produced over the lifetime of a specimen or sample. The identifier provides a public means of establishing long-time preservation that also classifies how data can be used, published or republished, and archived, which ultimately renders the original data and its derivatives more FAIR.

In the CT community, we face several obstacles concerning FAIR data management, especially regarding the use of PIDs. In the past five decades, little progress has been made to standardize CT metadata fields. As mentioned above, no universal non-proprietary output file formats exist that can be used interchangeably with a variety of CT manufacturers and segmentation software [10]. This absence of standardization can lead to data duplication or misinterpretation of important details such as scanning and reconstruction parameters, spatial axes, resolution and specimen attributes. While ORCIDs and DOIs are regularly used within the CT community, the use of PIDs is infrequent, confusing and disorganized. To address this, while also building on existing infrastructure, we recommend the adoption of a PID infrastructure incorporating the following well-established identifiers: RRIDs, RORs, ARKs and ORCIDs. This process will ensure that CT data are more discoverable, accessible, interoperable, reusable and machine-actionable [64]. In fact, such a PID infrastructure is ideally suited to the various needs of CT-data users, which consist of funders, research scientists, post-docs, laboratory technicians and managers, undergraduate and graduate students, publishers, archivists, data managers, repository representatives and software developers, industry vendors and manufacturers, engineers, policymakers and 3D-technology enthusiasts.

4.1. From specimen to archive: how PIDs mark the journey of a CT scan

The X-rays of a CT scanner are used to produce a series of 2D images with defined spatial relationships. These images are mathematically transformed into a secondary set of visualizations that represent cross-sectional slices through the object, stacked from top to bottom. Using this stack, which is a volumetrically accurate representation of the radiodensity of the original object, practitioners render 3D digital models. These models are, in turn, spatially accurate representations of the original object. They can be output into standard, shareable 3D surface file formats (e.g. .obj, .ply, .stl) with qualitative (e.g. presence or absence of specific features) and quantitative (e.g. size, shape) characteristics, including metadata annotations that detail sample provenance, scanning parameters and other relevant information [30]. These models are used for subsequent research-related tasks, such as taking measurements, running experiments and producing 3D prints. They are often archived for posterity, disseminated publicly and/or used in subsequent scientific studies. Best practices in CT research—from the initial scanning of a physical specimen to its archiving in a digital repository and eventual reuse—are marked by the generation and linkage of PIDs at each of these aforementioned steps. For instance, a natural history specimen might receive a PID from a national or institutional registry when it is first catalogued [14,45]. When the specimen is scanned, the resulting image dataset is assigned a separate PID. The institutional registry is recorded within the PID metadata, linking it back to the source specimen [45,65]. This linkage ensures that even as the data moves through various stages of processing and analysis, their provenance and related information remain intact and easily traceable. These linkages are the key to PIDs ensuring that data remains FAIR, which is achieved through several mechanisms (table 5).

4.2. PID integration connects publications, datasets and facilities

PIDs also play a crucial role in integrating metadata with datasets, enhancing their discoverability and reusability. For example, repositories such as iDigBio, MorphoBank and MorphoSource (see table 2) reference datasets using PIDs (e.g. DOI, Taxonomic IDs) to manage and share CT data. These repositories require that uploaded datasets be accompanied by comprehensive metadata and that both the data and metadata are assigned a PID. This practice facilitates the organization and retrieval of data within the repository and ensures that datasets can be linked to related publications, enabling a more integrated research environment. PID-enabled integration is similarly essential in connecting various elements of scientific productivity beyond the repository. For instance, a single PID can link a publication to its underlying dataset, ensuring that readers can directly access the associated data. This linkage also connects datasets to the facilities where the data were generated, such as specific CT scanning facilities.

By serving as the linkages within the research ecosystem, PIDs enhance the visibility and effect of research outputs. They ensure that datasets, publications and facilities are coherently interconnected, facilitating the reliable flow of information and enabling researchers to build on each other's work more effectively. Indeed, this is a cornerstone of how PIDs provide the associations needed to keep science open. By significantly enhancing the discoverability of data via stable and unique references that can be easily indexed and searched, PIDs improve the discoverability and reusability of CT scan data and derivatives. This discoverability and reusability is particularly important for large-scale data initiatives, where the large volume and diversity of data can otherwise present challenges for finding specific datasets [50]. For example, repositories such as the Open Science Framework [66] have adopted PIDs to ensure that large volumes of data can be consistently referenced and accessed.

The role of PIDs extends beyond enhancing discoverability; they also facilitate the connection among different research components. Many repositories employ PIDs to link datasets with publications, grants and other related research outputs, creating a rich network of interconnected information that supports data reuse and interdisciplinary research [67]. Networks, like NoCTURN [10], are using

Table 5. Mechanisms by which PIDs facilitate the implementation of FAIR principles by enhancing dataset interoperability across platforms.

FAIR principle	role of PIDs	key benefits	examples
findability	PIDs provide a unique and permanent reference for datasets, making them easily searchable and retrievable in digital repositories and search engines. They ensure consistent referencing in publications and metadata records.	enhances visibility and traceability, supports indexing in repositories and search engines.	PIDs combined with comprehensive metadata improve findability [50].
acessibility	PIDs link datasets to their metadata and repository entries, ensuring stable access over time, even if the repository location changes. They can also link to licensing and access information, promoting equitable access.	ensures long-term accessibility, stable access points despite repository changes, links to licensing/access information.	repositories using PIDs provide reliable access, as encouraged by the Nelson memorandum [23].
interoperability	PIDs standardize dataset identification across systems, allowing integration into various workflows, tools and platforms. They also enable datasets to be linked with other digital resources like software and computational models.	supports cross-disciplinary research, facilitates integration of datasets and tools, enhances collaboration across platforms.	standardization of PIDs promotes interoperability between datasets and digital resources [45].
reusability	PIDs maintain dataset integrity and provenance, linking data to detailed metadata that describe its origin and modifications. This linkage supports long-term reuse, citation and tracking of data usage and effect.	ensures dataset reuse, tracks data citations, supports replication of studies and building upon previous research.	PIDs enable accurate long-term data reuse and citation tracking [7].

PIDs to highlight facility capabilities and locations alongside the datasets they produce, allowing researchers to trace samples and scans back to their origins. These connections facilitate an understanding of the contexts in which the digital data were generated. Such transparency is essential for verifying the authenticity and reliability of data [50], important considerations when undertaking verification studies or replicating prior research results.

In essence, PIDs act as the glue connecting the various elements of scientific productivity and exemplify the goals of the 2022 Nelson memorandum. PIDs ensure that data remain linked to their metadata, facilitating accurate interpretation and reuse. PIDs connect publications to both the researchers who created them and the datasets they reference, promoting transparency and reproducibility. In addition, PIDs link datasets to the facilities where they were generated, providing context and enhancing trust in the data [3,45]. PIDs provide permanence or 'digital stickiness' (see [68]) enabling the research community to achieve and maintain greater openness, collaboration and scientific advancement [50].

4.3. PID challenges, limitations and future identifier perspectives

It is also important to highlight that the implementation of certain PID systems is incompletely realized. For example, many identifier types are excluded from the proposed US National PID strategy, such as those for images, licenses, methods and protocols. Similarly, PID systems that are still in early developmental stages or that have emerged as offshoots of already well-structured and funded PID projects like DOCID (Digital Object Container Identifier), RAiD (Research Activity Identifier), RRIDs and dARKs (a decentralized blockchain implementation of Archival Resource Keys; [69]) have not been embraced. Practical factors such as implementation costs, data security or risks, federalization versus centralization and ease of use also factor into PID design and implementation choices (e.g. [35,54,70]). Sometimes the scope of benefits may not become apparent until after implementation, but PIDs are now widely recognized for their value to archives, historical records, workflow and processing automation, and interoperable metadata exchange—among other applications. The value proposition can be weakened by broken links (ARKs), loss of funding for projects or organizations (RRIDs and RORs), human error and provider or host (organization, individual or project) neglect (e.g. not updating systems, not correcting broken links, using outdated versus real-time data).

While the challenges described above are certainly meaningful [71], new technologies that enhance insights across data production, analysis and distribution pipelines (e.g. machine learning) are excellent tools to aid the growing demands of data accessibility. Future improvements should focus on scalability, seamless integration with emerging technologies and the development of user-friendly, open systems to better serve the CT community. It is imperative that CT users—and practitioners of other data-heavy sciences—embrace comprehensive and standardized PID infrastructures to promote transparency, reproducibility and collaboration in alignment with inclusive processes. Open science is rooted in the principle of making research products available to all, while respecting diverse cultures and maintaining security, privacy and equity—values that the effective use of PIDs inherently supports.

5. Conclusion

In many ways, CT data have made progress towards openness: the creation of freely accessible virtual and physical resources in the form of data, and data derivatives are becoming more common and greater in number. These resources have led to the production of open software programmes, research data and publications, as well as educational resources that can be used within and outside of academic settings. However, making academic and non-clinical CT data fully FAIR and open will require expanding deliberate efforts through the establishment of proactive participation and implementation from the community. As a community, we see this as an important and timely narrative about what we require to establish data accessibility, how we aim to achieve this goal, and who is best positioned to implement and benefit from these innovations. The US implementation of open science infrastructure trails behind those of other regions, but concerted investment in interoperable metadata schema would help to close the gap. Crucially, metadata provides context to those who seek to understand the origin, methodology and limitations of previous research [72]. By including standardized descriptors and formats that enable different systems and tools to understand and work with the data, we can facilitate interoperability by fostering seamless integration and comparison of data from different sources [73]. This allows data from previous research to be interpreted and repurposed by users for new projects. Similarly, metadata helps support data preservation, including the history and context of data creation, modification and archiving—all vital to ensuring that data remain a usable, intelligible and reliable resource over time. Adopting such an interconnected system should be realized using trackable PIDs, linking metadata with data and its derivatives. PIDs provide the mechanism to ensure each element along the use and reuse chain are accounted for and linked. This common infrastructure is integral to the Nelson [23] vision because it will support global discoverability, interoperability, reusability, preservation, transparency and ethical or legal compliance of CT data [74]. Global scientific cooperation, thus, demands both participation and intention. NoCTURN envisions adaptable workflows emerging from this system that will promote openness not only for specialists but for all communities invested in education and outreach such as libraries, public schools, makerspaces, museums and non-profit organizations.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The data from the survey are available in the article's supplementary material [75].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. K.H.: conceptualization, project administration, supervision, writing—original draft, writing—review and editing; M.C.: funding acquisition, writing—original draft, writing—review and editing; C.C.: resources, writing—original draft, writing—review and editing; F.G.: visualization, writing—original draft, writing—review and editing; J.G.: data curation, visualization, writing—original draft, writing—review and editing; A.H.: writing—original draft, writing—review and editing; J.H.: writing—original draft, writing—review and editing; J.J.: writing—original draft, writing—review and editing; R.E.J.: writing—original draft, writing—review and editing; L.M.L.: visualization, writing—original draft, writing—review and editing; E.M.-S.: writing—original draft,

writing—review and editing; H.F.S.: writing—original draft, writing—review and editing; C.M.Z.: writing—review and editing; P.M.G.: conceptualization, funding acquisition, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The American Museum of Natural History, the Florida Museum of Natural History, the National Science Foundation (OAC-2226184 to M. Chase and P. Gignac, OAC-2226185 to E. Stanley and A. Krause, OAC-2226186 to J. Maisano), the University of Arizona, the University of Florida, the University of Texas at Austin.

Acknowledgements. We would like to thank Martin Halbert, NSF Science Advisor for Public Access, and Plato Smith, NSF Program Officer, and all the other inaugural FAIROS Research Coordination Networks for their dedication to advancing scientific disciplines. We thank the full NoCTURN community for their time and contributions to FAIR and open science. Communities are successful through the support of their members, and we would like to acknowledge the following individuals for their presence and thoughtful feedback throughout the development of this manuscript: Philip Bastians, Stephanie Baumgart, Laura Bento Da Costa, Francesco De Carlo, Kate Dobson, Catherine Early, Katie Franklin, Casey Holliday, Rich Ketcham, Arthur de Lima, Murat Maga, Kristin Marsh, Teresa Porri, Zach Randall, Akinobu Watanabe, Jay Warnett and Alex Waters. As strong supporters of PID use, the authors would also like to acknowledge the following facilities for their support with the manuscript: American Museum of Natural History Microscopy and Imaging Core Facility (RRID:SCR_024760), New York Institute of Technology NYIT Visualization Center Core Facility (RRID:SCR_024652) and Smithsonian Institution National Museum of Natural History Scientific Imaging Core Facility (RRID:SCR_025090). We thank Ed Stanley for his assistance with the survey data, and the editorial team and reviewers at Royal Society Open Science for their thoughtful and constructive feedback during the review process.

References

1. Hong NPC, Cozzini S, Genova F, Hoffman-Sommer M, Hooft R, Lembinen L, Marttila J, Teperek M. 2020 *Six recommendations for implementation of fair practice: by the fair in practice task force of the european open science cloud fair working group*. Brussels: European Commission. (doi:10.2777/986252)
2. G20 Hangzhou Summit. 2016 2016 Hangzhou Summit: G20 2016 Innovation Action Plan. See https://www.g20.utoronto.ca/2016/160905-innovation.html (accessed 3 February 2025).
3. UNESCO. 2021 UNESCO recommendation on open science. In *41st Session of The General Conf. of the United Nations Educational, Scientific and Cultural Organization*. Paris, France: FR. (doi:10.54677/MNMH8546)
4. Schuster D, Friedman M. 2023 AMS publications support for open, transparent, and equitable research. *Artif. Intell. Earth Syst.* **2**, e230079. (doi:10.1175/jcli-d-23-0552.1)
5. Wilkinson MD *et al*. 2016 The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018. (doi:10.1038/sdata.2016.18)
6. Bicer T, Gürsoy D, Andrade VD, Kettimuthu R, Scullin W, Carlo FD, Foster IT. 2017 Trace: a high-throughput tomographic reconstruction engine for large-scale datasets. *Adv. Struct. Chem. Imaging* **3**, 6. (doi:10.1186/s40679-017-0040-7)
7. Blackburn DC *et al*. 2024 Increasing the impact of vertebrate scientific collections through 3D imaging: the openVertebrate (oVert) thematic collections network. *Bioscience* **74**, 169–186. (doi:10.1093/biosci/biad120)
8. Cunningham JA, Rahman IA, Lautenschlager S, Rayfield EJ, Donoghue PCJ. 2014 A virtual world of paleontology. *Trends Ecol. Evol*. **29**, 347–357. (doi:10.1016/j.tree.2014.04.004)
9. Faulwetter S. 2015 First steps towards the development of an integrated metadata management system for biodiversity-related micro-CT datasets. *Bruker microct user meet*.
10. Gignac PM *et al*. 2024 The role of networks to overcome large-scale challenges in tomography: the non-clinical tomography users research network. *Tomogr. Mater. Struct*. **5**, 100031. (doi:10.1016/j.tmater.2024.100031)
11. Hardesty J, Johnson J, Wittenberg J, Hall N, Cook M, Lischer-Katz Z, Xie Z, McDonald R. 2020 3D Data repository features, best practices, and implications for preservation models: findings from a national forum. *Coll. Res. Libr*. 789. (doi:10.5860/crl.81.5.789)
12. Kuzminsky SC, Gardiner MS. 2012 Three-dimensional laser scanning: potential uses for museum conservation and scientific research. *J. Archaeol. Sci*. **39**, 2744–2751. (doi:10.1016/j.jas.2012.04.020)
13. Prior F, Almeida J, Kathiravelu P, Kurc T, Smith K, Fitzgerald TJ, Saltz J. 2020 Open access image repositories: high-quality data to enable machine learning research. *Clin. Radiol*. **75**, 7–12. (doi:10.1016/j.crad.2019.04.002)
14. Boyer DM, Gunnell GF, Kaufman S, McGeary TM. 2016 Morphosource: archiving and sharing 3-D digital specimen data. *Paleontol. Soc. Pap*. **22**, 157–181. (doi:10.1017/scs.2017.13)
15. Zhou Y, Zhang Z, Wang X, Sheng Q, Zhao R. 2024 Multimodal archive resources organization based on deep learning: a prospective framework. *AJIM* **77**, 530–553. (doi:10.1108/AJIM-07-2023-0239)
16. Buhr CC, Kleiner T. 2012 European open data policy: challenges and opportunities. *Zeitschrift Für Politikberatung (ZPB)/Policy Advice and Political Consulting* **5**, 141–146.

17. European Commission. 2012 Commission recommendations on access to and preservation of scientific information. *Off. J. Eur. Union* **L194**, 39–43.
18. European Commission. 2024 Open access - European Commission. See https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/open-access_en (accessed 21 March 2025).
19. European Commission. 2024 Open Science - European Commission. See https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science_en (accessed 21 March 2025).
20. Rylance R et al. 2016 *Concordat on Open Research Data*. UK Research and Innovation. See https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf.
21. 2018 *Realising the potential: Final report of the Open Research Data Task Force*. Open Research Data Task Force.
22. Holdren JP. 2013 Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research
23. Nelson A. 2022 Ensuring free, immediate, and equitable access to federally funded research. Office of science and technology policy. See https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf.
24. El-Sherbini M, Klim G. 2004 Metadata and cataloging practices. *Electron. Libr*. **22**, 238–248. (doi:10.1108/02640470410541633)
25. Habermann T. 2020 Metadata and reuse: antidotes to information entropy. *Patterns* **1**, 100004. (doi:10.1016/j.patter.2020.100004)
26. Woolcott L. 2017 Understanding metadata: what is metadata, and what is it for? *Iii* **45**, 669–670. (doi:10.1080/01639374.2017.1358232)
27. Sarkans U *et al*. 2021 REMBI: recommended metadata for biological Images—enabling reuse of microscopy data in biology. *Nat. Methods* **18**, 1418–1422. (doi:10.1038/s41592-021-01166-8)
28. Keklikoglou K, Faulwetter S, Chatzinikolaou E, Wils P, Brecko J, Kvaček J, Metscher B, Arvanitidis C. 2019 Micro-computed tomography for natural history specimens: a handbook of best practice protocols. *Eur. J. Taxon* **522**, 1–55. (doi:10.5852/ejt.2019.522)
29. Jeffrey K, Koskela R. 2014 RDA Metadata principles and their use. *Research (Lambertville).* **512**, 6.
30. Balk MA *et al*. 2024 A FAIR and modular image-based workflow for knowledge discovery in the emerging field of imageomics. *Methods Ecol. Evol.* **15**, 1129–1145. (doi:10.1111/2041-210x.14327)
31. Chatzinikolaou E, Keklikoglou K, Grigoriou P. 2021 Morphological properties of gastropod shells in a warmer and more acidic future ocean using 3d micro-computed tomography. *Front. Mar. Sci*. **8**, 645660. (doi:10.3389/fmars.2021.645660)
32. Ewert U, Fuchs T. 2017 Progress in digital industrial radiology. Pt. 2, Computed tomography (CT). *Badania Nieniszcz. Diagn* 7–14. (doi:10.26357/BNiD.2017.018)
33. Jiřík M *et al*. 2018 Generating standardized image data for testing and calibrating quantification of volumes, surfaces, lengths, and object counts in fibrous and porous materials using X-ray microtomography. *Microsc. Res. Tech*. **81**, 551–568. (doi:10.1002/jemt.23011)
34. Stock SR, De Carlo F. 2023 Meta-data for absorption tomography measurements. *Tomogr. Mater. Struct*. **3**, 100015. (doi:10.1016/j.tmater.2023.100015)
35. Johnston LR *et al*. 2024 Seek and you may (not) find: a multi-institutional analysis of where research data are shared. *PLoS One* **19**, e0302426. (doi:10.1371/journal.pone.0302426)
36. Cnudde V, Boone MN. 2013 High-resolution X-ray computed tomography in geosciences: a review of the current technology and applications. *Earth Sci. Rev*. **123**, 1–17. (doi:10.1016/j.earscirev.2013.04.003)
37. Wildenschild D, Sheppard AP. 2013 X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Adv. Water Resour*. **51**, 217–246. (doi:10.1016/j.advwatres.2012.07.018)
38. Kelly B, Closier A, Hiom D. 2005 Gateway standardization: a quality assurance framework for metadata. *Libr. Trends* **53**, 637–650.
39. Feldkamp LA, Davis LC, Kress JW. 1984 Practical cone-beam algorithm. *J. Opt. Soc. Am. A* **1**, 612–619. (doi:10.1364/josaa.1.000612)
40. Davies TG *et al*. 2017 Open data and digital morphology. *Proc. R. Soc. B Biol. Sci.* **284**, 20170194. (doi:10.1098/rspb.2017.0194)
41. Klump J, Huber R. 2017 20 years of persistent identifiers – which systems are here to stay? *Data Sci. J.* **16**, 9. (doi:10.5334/dsj-2017-009)
42. Blake M *et al*. 2022 Data Curation Network: CURATED Training.
43. Brown J, Jones P, Meadows A, Murphy F, Clayton P. 2021 UK PID consortium: cost-benefit analysis (doi:10.5281/zenodo.4772627)
44. Freitas A, Curry E. 2016 Big data curation. In *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in europe* (eds JM Cavanillas, E Curry, W Wahlster), pp. 87–118. Springer. (doi:10.1007/978-3-319-21569-3_6)
45. Moore J, Rountrey A, Kettler HS. 2022 3D data creation to curation: community standards for 3D data preservation. *Assoc. Coll. Res. Libr* **45**, 223–226. (doi:10.11588/ai.2022.1.95349)
46. McMurry JA *et al*. 2017 Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biol.* **15**, e2001414. (doi:10.1371/journal.pbio.2001414)
47. Australian Research Data Commons. 2024 *Australian national Persistent Identifier (PID) strategy 2024*.
48. Barsky E. 2020 *Persistent identifiers in canada: white paper*. (doi:10.14288/1.0395014)
49. Brown J. 2023 Developing a persistent identifier roadmap for open access to UK research. *Open Scholarsh*
50. ORFG P.I.D. Strategy Working Group. 2024 Developing a US National PID Strategy. *Univ. Calif. Curation Cent UC3* (doi:10.5281/zenodo.10811008)
51. Brown C, Simons N, Bangert D, Sadler S. 2023 *RDA national PID strategies guide and checklist*. Zenodo. See https://doi.org/10.15497/RDA/00091.
52. Meadows A, Barilla MB. 2024 *The Scholarly Kitchen: What's Hot and Cooking inScholarly Publishing*. #PIDfest – Bringing the PID Community Back Together Again!
53. Meadows A. 2024 A successful start to a new festival of identifiers: PIDfest 2024. In *The scholarly kitchen: what's hot and cooking in scholarly publishing*.

54. Meadows A. 2023 The scholarly kitchen: what's hot and cooking in scholarly publishing. In *The orcid us consortium at five: what's worked, what hasn't, and why?* Society for Scholarly Publishing. See https://scholarlykitchen.sspnet.org/2023/06/21/the-orcid-us-consortium-at-five-whats-worked-what-hasnt-and-why/.
55. Meadows A. 2023 The scholarly kitchen: what's hot and cooking in scholarly publishing. In *Why pid strategies are having a moment — and why you should care*. Society for Scholarly Publishing. See https://scholarlykitchen.sspnet.org/2023/01/25/why-pid-strategies-are-having-a-moment-and-why-you-should-care/.
56. Chiarelli A. 2022 *The Confused Developer: URL, URI, URN: What's the Difference*. Auto0 by okta.
57. Meadows A. 2017 *Ten reasons to get—and use—an ORCID iD*. See https://www. elsevier. com/connect/authors-update/ten-reasons-to-get-and-use-anorcid-id.
58. Meadows A. 2019 The scholarly kitchen. In *Are you ready to ror? an inside look at this new organization identifier registry*. Society for Scholarly Publishing. See https://scholarlykitchen.sspnet.org/2019/12/04/are-you-ready-to-ror-an-inside-look-at-this-new-organization-identifier-registry/.
59. Mayernik M, Johnson A, Julian R, Murray M, Mundoma C, Ranganath A, Stossmeister G. 2024 Persistent identifiers for instruments and facilities: current state, challenges, and opportunities. *J. EScience Librariansh* **13**, e964. (doi:10.7191/jeslib.964)
60. Smith AM, Katz DS, Niemeyer KE, Group FSCW. 2016 Software citation principles. *PeerJ Comput. Sci.* **2**, e86. (doi:10.7717/peerj-cs.86)
61. Chue Hong NP *et al*. 2022 FAIR principles for research software (FAIR4RS principles). *Zenodo* (doi:10.15497/RDA00068)
62. Cosmo RD, Gruenpeter M, Zacchiroli S. 2020 Referencing source code artifacts: a separate concern in software citation. *Comput. Sci. Eng*. **22**, 33–43. (doi:10.1109/mcse.2019.2963148)
63. Cosmo R, Gruenpeter M, Zacchiroli S. 2018 Identifiers for digital objects: the case of software source code preservation. In *Proc. of iPRES 2018: 15th Int. Conf. on Digital Preservation*, p. 9. Boston, MA, USA: iPRES.
64. Kim S. 2023 Introduction to pids: what they are and how to use them. In *Becker blog* (ed. K Seonyoung). St. Louis, MO: Bernard Becker Medical Library. See https://becker.wustl.edu/news/introduction-to-pids-what-they-are-and-how-to-use-them/.
65. Miralles A *et al*. 2020 Repositories for taxonomic data: where we are and what is missing. *Syst. Biol*. **69**, 1231–1253. (doi:10.1093/sysbio/syaa026)
66. Foster ED, Deardorff A. 2017 Open science framework (OSF). *J. Med. Libr. Assoc.* **105**, 203. (doi:10.5195/jmla.2017.88)
67. Haris M, Auer S, Stocker M. 2023 Scholarly knowledge graph construction from published software packages. In *Leveraging generative intelligence in digital libraries: towards human-machine collaboration* (eds DH Goh, SJ Chen, S Tuarob), pp. 170–179. Singapore: Springer Nature. (doi:10.1007/978-981-99-8088-8_15)
68. Kunze J, Calvert S, DeBarry J, Hanlon M, Janée G, Sweat S. 2016 Persistence statements: describing digital stickiness
69. Segundo W, Matas L, Nóbrega T, Filho JES, Mena-Chalco J. 2023 dARK: a decentralized blockchain implementation of ARK Persistent Identifiers. *BiblioCanto* **9**, 147–158. (doi:10.21680/2447-7842.2023v9n2ID33759)
70. Klausen MH. 2017 Even minor integrations can deliver great value – a case study. *Procedia Comput. Sci*. **106**, 153–159. (doi:10.1016/j.procs.2017.03.011)
71. Borgman CL, Groth P. 2025 From data creator to data reuser: distance matters. *Harv. Data Sci. Rev* **7**, 1–41. (doi:10.1162/99608f92.35d32cfc)
72. Hardwicke TE *et al*. 2018 Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *R. Soc. Open Sci.* **5**, 180448. (doi:10.31222/osf.io/39cfb)
73. Stocker M, Prinz M, Rostami F, Kempf T. 2019 Towards research infrastructures that curate scientific information: A use case in life sciences. In *Data Integration in the Life Sciences: 13th International Conference, DILS 2018*, pp. 61–74. Hannover, Germany: Springer International Publishing. (doi:10.1007/978-3-030-06016-9_6)
74. Fisher M, Fradley M, Flohr P, Rouhani B, Simi F. 2021 Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project. *Archaeol. Prospect*. **28**, 279–292. (doi:10.1002/arp.1816)
75. Hurdle K, Chase M, Crawford C, Goetz F, Gray J, Hall A *et al*. 2025 Supplementary material from: A FAIR Narrative on Open Science: When, Who, What, and How of Computed Tomography and the NoCTURN Network. FigShare. (doi:10.6084/m9.figshare.c.8192660)