Computational Humanities Research

www.cambridge.org/chr

Enabling reproducible research in past networks: A data documentation framework

Petra Heřmánková[1], Tom Brughmans[1], Matteo Mazzamurro[2], Peter Bjerregaard Vahlstrup[3] and Per Møldrup-Dalum[3]

[1]Classical Archaeology and Social Resilience Lab, Aarhus Universitet, Denmark; [2]School of Physical and Chemical Sciences, Queen Mary University of London, UK and [3]Center for Humanities Computing, Aarhus Universitet, Denmark

Research Article

Cite this article: Heřmánková Petra, Tom Brughmans, Matteo Mazzamurro, Peter Bjerregaard Vahlstrup and Per Møldrup-Dalum. 2026. "Enabling reproducible research in past networks: A data documentation framework." *Computational Humanities Research*, **2**:e29, https://doi.org/10.1017/chr.2026.10051

Received: 21 April 2026 UTC
Revised: 28 August 2026 UTC
Accepted: 22 September 2026 UTC

Keywords:
FAIR and LOUD data standards; historical sciences; past network research; reproducibility

Corresponding author:
Petra Heřmánková;
Email: petra.hermankova@cas.au.dk

Abstract

This article addresses a significant challenge hindering large-scale research into past human social networks: the lack of structured and standardised documentation for published network data, alongside inconsistent community-wide adoption of such practices. Although the community of practitioners applying network analysis to historical and archaeological sources remains small and the corpus of available network datasets is still limited, there is a clear demand for shared infrastructural resources, which are increasingly viewed as essential to sustaining and scaling the field. To overcome these obstacles, the authors propose a formulation of past network data standards that reflect current usage and identify missing components to achieve best practices while adhering to the FAIR and LOUD principles of Open Science. The proposed standards aim to enhance understanding of data provenance, licensing, and feasibility for specific research scenarios, while also promoting reproducibility and facilitating the automation of data reuse for large-scale comparative studies. Additionally, the *Past Networks Repository* was developed to support the uptake of these common standards and improve resource findability. The repository explicitly implements these data standards to facilitate collaboration and research dissemination.

Plain language summary

Past network research examines how people in the past were connected through friendship, family, trade, religion, politics, or the objects they made and used. Researchers reconstruct these connections and their development over time as maps of relationships, drawing on historical sources such as inscriptions, letters, administrative records, or archaeological assemblages. Over recent decades, many such network datasets have been created, but they have been published in inconsistent ways, stored in different file formats, with variable levels of documentation, and with few shared rules about how data should be described, licensed, or shared. This makes it very difficult for other researchers to find, trust, reuse, or compare these datasets, particularly when working at a large scale across multiple historical periods or regions.

This article tackles that problem directly and provides a tangible solution to the most pressing issues. The authors reviewed how past network datasets are currently documented and identified what is missing or inconsistent across the field. Drawing on widely adopted open science principles, specifically the FAIR (Findable, Accessible, Interoperable, and Reusable) and LOUD (Linked Open Usable Data) principles, they propose a set of practical documentation standards, distributed over 55 attributes, for how researchers should document and share network data derived from historical sources. These standards address key questions that any potential data user needs to answer before reusing someone else's dataset: *Where did the data come from? Under what conditions can it be used? Is it suitable for a particular research question?*

Crucially, the authors do not just propose standards in the abstract; they built a tool to put them into practice. The *Past Networks Repository* is an openly accessible online resource where researchers can deposit, find, and reuse historical network datasets that conform to these shared standards. By bringing data together in one place with consistent documentation, the repository makes it easier for researchers to build on each other's work, run large-scale comparative studies across different times and places, and reproduce published findings.

The ultimate goal is to make the study of past social networks more collaborative and reproducible and provide an infrastructure enabling future growth of the discipline: a field where individual research efforts contribute to a shared, growing body of high-quality, well-documented data rather than remaining isolated and difficult to access.

Introduction

The volume of published network data about the human past has experienced significant growth in recent times (Brughmans and Peeples 2023; Brughmans et al. 2023; Collar 2022; Graham,

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

Milligan, and Weingart 2015). This surge in data availability, however, is accompanied by a major obstacle hindering large-scale research of past phenomena: the lack of structured and standardised data documentation, as well as the community-wide adoption of such practices (Middle 2023). Without consistent documentation standards, the reproducibility of computational analyses, that is, the ability to re-run the same workflow on the same data and obtain identical results, and the replicability of broader findings, that is, the ability to apply equivalent methods to independent datasets and reach comparable conclusions, remain severely compromised. A recent survey of the community revealed that data-related and network construction issues topped the list of challenges practitioners experience in the field (Weiß et al. 2025). Only a third of the respondents had experience with open data publication, and "Only a minority of scholars were able to follow guidelines or find orientation concerning the best practices for publishing data, code and results online" (Weiß et al. 2025).

For past network research, the situation is attributable to several factors. Despite long-standing standardisation efforts in the Digital Humanities, exemplified by initiatives such as the Text Encoding Initiative (TEI) (TEI Consortium 2026) or the CIDOC Conceptual Reference Model (CIDOC CRM) (Bekiari et al. 2024), past network data lack dedicated documentation frameworks, in part due to its cross-disciplinary scope. This absence is further compounded by a shortage of repositories specifically tailored to the historical network data, as well as hesitant attitudes towards open science within certain segments of the community (Marwick 2017; Marwick et al. 2017; Nicholson et al. 2023; Olsson, Sköld, and Andersson 2024).

To address these challenges and prevent further exacerbation, we formulate past network data standards that reflect current usage, identify any missing components, and expand these standards to attain the highest level of best practice while adhering to the FAIR (Findable, Accessible, Interoperable, and Reusable) and LOUD (Linked Open Usable Data) data principles (Laguna-Palma and Barruezo-Vaquero 2024; Nicholson et al. 2023; Thiery 2019; Wilkinson et al. 2016). The existence of documentation standards facilitates our understanding of the data's provenance, licensing, and feasibility for specific research scenarios and alerts us to potential data flaws. Furthermore, standardised documentation promotes reproducibility and replicability while facilitating the automation of data reuse processes. This opens new avenues for large-scale comparative and longitudinal research, allowing for innovative methodologies and insights. To encourage widespread adoption, we maintain simplicity in the minimum documentation requirements while allowing for extensibility to optimise data accessibility.

To support the uptake of common standards, enhance resource discoverability, and establish a sustainable infrastructure for future growth, we have developed a new open repository, the *Past Networks Repository* (Brughmans et al. 2025). It provides a platform to document past network datasets deposited elsewhere using the newly established standards, thereby operationalising them through a dedicated, practical outlet.

This article is intended for researchers working with network data for the study of the human past, students and scholars who are collecting and publishing such data, as well as interested individuals from related disciplines seeking inspiration or a deeper understanding of how past network data is structured and managed.

Community stocktaking

Authors of this article conducted a thorough stocktaking among past network research community members and practitioners to gain an overview of the current landscape of challenges in the field. This scoping exercise took place in January 2023 during a workshop at Aarhus University (Denmark) and later that year at the Santa Fe Institute (United States). Over 30 researchers and practitioners identified two issues that significantly hinder both the research process and collaboration within and beyond the community: limited practice of structured data documentation and the absence of documentation standards. The results of the stocktaking were further confirmed by a recent online survey conducted among the community, which focused on the reflections of 129 respondents on the current practices and methods in historical network research (Weiß et al. 2025).

Cultural shift

Reproducibility has become a contested norm across academic disciplines. While it is increasingly required in computational and natural sciences, its application to humanities research remains debated: humanistic inquiry is seen as interpretive, context-dependent, qualitative assessment-driven, and typically small-scale in scope, and resistant to the notion that analyses can or should be independently replicated (Joyeux-Prunel 2024; Nicholson et al. 2023). Nevertheless, computationally intensive humanities research, such as past network analysis, occupies a middle ground: full replication in the natural-science sense may not be feasible given the fragmentary nature of historical data, but the community should strive for computational reproducibility. Without it, results remain isolated and difficult to evaluate, compare, or build upon, and compromise the cumulative potential of the field (Burrows 2023; Sobotkova and Hermankova 2022). The ability to rerun documented code and apply consistent methodology across geographically and temporally diverse datasets enables systematic investigation of large-scale social phenomena, such as the development of urbanism, patterns of human mobility, or the diffusion of ideas across regions and periods or evolving the structure of the society (Brughmans et al. 2023; Collar 2022).

Achieving this, however, requires overcoming two interrelated challenges. The first challenge is primarily cultural and extends beyond practitioners of past network research (Nicholson et al. 2023; Thiery 2019). There is limited practice in documenting the processes for network dataset creation and analysis in a manner that enables replication (Weiß et al. 2025). This issue can be gradually addressed, particularly once the technical obstacles and standard formulations of the second challenge are resolved.

While methods and coding languages are often mentioned in academic publications, the specific software implementations, libraries used, and transformations applied to the input data are frequently omitted from both the main text and supplementary materials. As a result, readers are unable to replicate the analysis, verify critical steps in the data analysis pipeline, or reuse relevant portions of the code. Data analysis typically involves a series of decisions and tasks that significantly shape both the dataset and the analytical outcomes. As a community, we should collectively encourage and support researchers in publishing their code alongside their data in open and accessible formats, such as Jupyter or Quarto Notebooks, to foster transparency, reproducibility, and reusability. Best practice recommendations should, however, go beyond

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

merely sharing code and data. They need to include comprehensive documentation: well-annotated code, detailed READ.ME files, a summary of software requirements, specification of libraries and packages used, minimal software versions, and, where applicable, virtual or containerised environments and implementation of version control best practices (Huskey 2024; Marwick 2017). Many journals with a computational focus already require this as a standard practice, and high-quality documentation is recognised as the key factor in (re)usability and reproducibility in Digital Humanities and beyond (Middle 2023). However, more traditional humanities journals lag behind in this regard. A further crucial role is played by educational initiatives, such as creating teaching/learning resources, organising networking events, and presenting case studies that illustrate the practical benefits of data documentation. *The Barcelona Past Networks Summer School* (Mazzamurro et al. 2026) offers an example, where the community actively promotes best practices each year among a group of new practitioners and learners.

Technical barriers

The second identified group of challenges is more technical and more readily solvable. It concerns conceptual and semantic variability, as well as the absence of minimally agreed-upon standards (Middle 2023).

The development of universal standards has been a core focus across the Digital Humanities for the past decades. Built to solve the fragmentation of siloed collections and enable computational analysis on complex cultural data, foundational frameworks like CIDOC CRM (Bekiari et al. 2024; CIDOC CRM Special Interest Group 2006), Machine-Readable Cataloguing (MARC) (Library of Congress 1999), and Dublin Core (Dublin Core Metadata Initiative 1995) established the core infrastructure for heritage information management and modelling, bibliographic discovery, and resource description, while the TEI (TEI Consortium 1987) provided a rich XML schema for structural text encoding. As specialised subfields emerged, whose datasets frequently underpin past network science research, scholars adapted these foundational frameworks into domain-specific extensions to capture localised nuances. For instance, EpiDoc XML adapted TEI for epigraphy (Elliott et al. 2006–2026), while semantic initiatives like the ARCH-ON Project (Deckers et al. 2026) developed standardised infrastructure for archaeological object classification, the Sampo model (Hyvönen 2023) for cultural heritage semantic portals, CIDOC CRMtex (Murano, Felicetti, and Fafalios 2023) for ancient text modelling, and EpOnt (Bodard et al. 2021; Prag and Asif 2026) for epigraphic ontologies and controlled vocabularies (Heřmánková et al. 2025) – while leveraging the Linked Open Data model (Berners-Lee 2006) to bridge disciplinary siloes. However, because network research spans diverse pathways from the Humanities to the Natural Sciences, standardisation remains a complex, ongoing process balancing disciplinary requirements with unified interoperability.

A review of the existing projects and network data repositories, such as the Colorado Index of Complex Network (ICON) (Clauset, Tucker, and Sainz 2016), Stanford Large Network Dataset Collection (Leskovec and Sosič 2016), and The SuiteSparse Matrix Collection (Davis and Hu 2011), revealed the set of commonly used attributes and their protocols, but also noted high semantic variability as well as uneven level of granularity of the provided metadata. Moreover, the licensing information was not always included, potentially hindering data reuse and not adhering to the basics of the FAIR and LOUD data principles. We have compiled a list of commonly used general-purpose and domain-specific repositories for data and code sharing, with the source of their funding, as shown in Table 1. Currently, the discoverability of past network research and related data remains a challenge, severely limiting their potential reusability. Datasets and related analyses are scattered across various outlets, ranging from specialised network publications and collections to general disciplinary journals and repositories, making initial discovery time-consuming and inconsistent.

Community-defined scenarios

To design a solution aligned with current and emerging practices, we further invited community members at the Aarhus workshop to define user scenarios that reflect both actual and ideal workflows. These community-defined scenarios highlight the diverse use cases envisioned for the *Past Networks Repository* and emphasise the importance of interoperability, documentation, and openness.

Participants identified the wish to reproduce existing analyses by accessing the original datasets, whether hosted in the repository or linked from associated publications. To support comparative studies, the repository must provide access to standardised datasets and metadata, enabling scholars to identify broader patterns, test for structural similarities, and evaluate how unique or representative their networks are in different regional or thematic contexts. Researchers may wish to collate disparate datasets, potentially separated by time, geography, or diverging conceptual and structural conventions, into larger-scale structures. This requires procedures for reconciling differing data standards and their alignment. In addition, a repository should support innovation, allowing researchers to repurpose existing datasets to explore new questions not originally anticipated. This, in turn, relies on rich and precise metadata that facilitates reinterpretation and creative reuse. Some intended scenarios involve adapting or expanding existing datasets, which requires access to raw data to assess the original modelling choices and compare alternative interpretations, as well as clearly defined licensing information.

The members of the community agreed that the technical solution has to remain simple, relying on straightforward aggregation of metadata indexes. When depositing data or code, authors must supply a basic metadata set, specify the repository storing the raw data and/or network in a non-proprietary format, and list the parameters used to generate the network. This metadata needs to be detailed enough to allow others to replicate the methods and understand the underlying decisions, yet remain easy to use. To support this, depositors can emphasise key elements and present them in a user-friendly way. This approach allows for flexibility in defining the network according to the user's specific needs, rather than being constrained by the original curator's interpretation (Sobotkova and Hermankova 2022).

In an ideal scenario, each repository entry includes key descriptive metadata and basic network properties, such as the number of nodes and edges, average degree, and network density. This information provides researchers with an initial overview of the dataset, helping them assess its relevance and inform their decision-making process (Middle 2023). To further support the discoverability of the research and data, authors need to be able to assign a set of tags from a controlled thesaurus, ideally maintained by a dedicated working group of practitioners. These tags are supplemented by author-defined keywords to provide

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

Table 1. Examples of existing repositories and their disclosed sources of funding

Repository	Governance	Affiliation/Funding
General-purpose science repositories		
Zenodo	Non-profit	European Union via CERN
Open Science Framework	Non-profit	Federal agencies and private foundations
The Dataverse Project	Institutional	Harvard Institute for Quantitative Social Science; institutional support
Figshare	Commercial	Digital Science (Springer Nature subsidiary)
Domain-specific repositories for archaeological data		
Open Context	Non-profit	Alexandria Archive Institute; NEH, NSF
tDAR	Non-profit	Arizona State University; Mellon Foundation, NSF, NEH
ADS	Non-profit	University of York; UKRI/AHRC and JISC
DANS	National infrastructure	Dutch KNAW and NWO
iDAI	Governmental	German Archaeological Institute; Federal Foreign Office
ARIADNE+	Research infrastructure	EU Horizon 2020
Network-specific repositories		
Sparse Matrix Collection	Academic	Texas A&M University; NSF and University of Florida
SNAP	Academic	Stanford University; DARPA, NIH, NSF, Stanford Data Science Initiative
ICON	Academic	University of Colorado Boulder
Thematic dataset collection on covert networks	Academic	Mitchell Centre for Social Network Analysis, University of Manchester, Leverhulme Trust
Netzschleuder	Academic	Open collection by Tiago P. Peixoto
Code-sharing repositories		
GitLab	Open-source; Self-hostable	Community-developed
GitHub	Commercial	Microsoft
Bitbucket	Commercial	Atlassian
Codeberg	Non-profit	Donations and membership fees
Software Heritage	Non-profit	UNESCO; various sponsors

Note: More extensive information is available in Nicholson et al. (2023).

additional specificity. An autofill feature suggests keywords based on previous entries. Additionally, creators of the repository entry are able to provide an unconstrained, detailed description to guide potential users on the possibilities and limitations of the dataset, facilitating its effective reuse. Over time, once a critical amount of data is collected, the community-based working group should then review and refine the keyword set to promote consistency and improve standardisation across records.

Minimal documentation requirements

Clear, comprehensive documentation of minimal data standards is essential for reproducibility in Digital Humanities and a prerequisite for the usability of digital resources and transferability of methods (Middle 2023). Drawing on community input, we defined a set of minimal documentation requirements and developed a foundational yet adaptable protocol for past network data. This protocol is intended as a practical guide for practitioners, students, and researchers when creating and depositing new entries, and encourages revision and enhancement of existing resources. In practice, it directly shaped the design and structure of the *Past Networks Repository*, as discussed below.

We identified 55 attributes reflecting best practice and grouped them into several categories by their thematic focus, such as essential identification and authorship, or semantic and technical descriptions of the network and its construction process. Of these attributes, 27 are mandatory and form the best practice minimal documentation requirements, while the remaining 28 are optional based on recommendations from the community. This attribute selection balances the need for presenting essential information with practical considerations of what researchers can reasonably be expected to provide during the documentation process. The groups are:

1. Essential identification and authorship information: An overview of where the network data is deposited and published, and details of how to access it (22 attributes, 13 mandatory).
2. Network topic: The network's topic, time period, geographical area, and description of uncertainty (13 attributes, six mandatory).
3. Technical parameters: Quantitative descriptors of network structure and network size (14 attributes, four mandatory).
4. Network construction and provenance: An overview of the data curation process and sources (six attributes, four mandatory).

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

Table 2. Essential identification and authorship information attributes required for network dataset documentation

Attribute name	Type	Description	Mandatory
Name	Text	The name of the network	Yes
Network ID	Numeric	Unique identifier, ideally automatically generated and computer-actionable	Yes
Data URL	Hyperlink	URL linking to the location where the network can be accessed and downloaded	Yes
Canonical Citation of the Network Dataset	Text	Canonical citation of the network dataset	Yes
Description	Text	Detailed description of the network	Yes
File Format	Controlled vocabulary, Multiple	File format(s) of the network dataset, for example, graphJSON, matrix, graphml; multiple formats allowed	Yes
Authors	Child entity, Multiple	Authors and contributors who created the network; at least one author required	Yes
— *First Name*	Text	First name of the author/contributor	Yes
— *Last Name*	Text	Last name of the author/contributor	Yes
— *Affiliation*	Text	Affiliation of the author/contributor	No
— *ORCID ID*	Text	ORCID identifier of the author/contributor	No
License	Controlled vocabulary	License under which the dataset is published; use "Restricted access" if unavailable	Yes
Tags	Text, Multiple	Keywords describing the network	Yes
Collection Tags	Text, Multiple	Collection name(s) for grouping networks, for example, projects or publications	No
Image	Image	Visual representation of the network data	No
Funding	Text	Information about funding sources	No
Data Publisher	Text	Repository or entity that published the dataset	No
Related Publications	Child entity, Multiple	Information about related publications	No
— *Canonical Citation of Reference*	Text	Canonical citation of the related publication reference	No
— *Reference URL*	Hyperlink	Hyperlink to the reference; DOI URL preferred if available	No
Record Creation Date	Timestamp	Date when the record was created; ideally automatically assigned	Yes
Record Last Updated	Timestamp	Date of the last update to the record; ideally automatically assigned	Yes

Note: The 22 attributes range from basic identification (name, ID, and data URL) to authorship details and licensing information. Of these, 13 are mandatory to ensure minimal documentation standards and enhance discoverability, while nine remain optional to provide additional context.

Essential identification and authorship information

Network datasets require systematic documentation to ensure their discoverability, reusability, and correct attribution within the research community. The group of 22 essential identification attributes enables researchers to locate and access relevant datasets through searchable metadata, provide essential context about the network's construction and limitations, ensure proper citation and credit to original creators of the data and published research, and facilitate reproducible research by documenting technical specifications and methodological details, all in line with FAIR data principles and Open Science best practice, listed in Table 2. Of these attributes, 13 are mandatory to ensure minimal documentation standards and enhance discoverability, while nine remain optional to provide additional context. The mandatory attributes provide a basic means for the identification of the network, its deposition location, format, authorship, and access to the published data. The optional attributes provide details about related projects, publications, and funding sources, as well as topics covered by the data, to improve the findability of recorded datasets.

Network topic

Network datasets contain rich semantic information that defines their topical scope and temporal boundaries. Understanding the meaning and context of network elements, that is, what the nodes and edges represent, when and where the network existed, and what additional attributes are captured, is crucial for researchers to assess the dataset's relevance to their research questions and to interpret findings appropriately.

The 13 semantic attributes in Table 3 provide essential context about the network's subject matter, geographical and temporal scope, and the types of relationships and entities it captures. Of these 13 attributes, six are mandatory to ensure basic semantic documentation, while seven are optional attributes that provide enhanced detail about node and edge characteristics, including

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

Table 3. The semantic attributes describe the topical and contextual dimensions of network datasets, covering temporal scope, geographical provenance, and the meaning of network components

Attribute name	Type	Description	Mandatory
Year From	Date/numeric	Starting date of the time period (YYYY format; for BCE dates use a negative value; 0 is not permitted)	Yes
Year To	Date/numeric	Ending date of the time period (YYYY format; for BCE dates use a negative value; 0 is not permitted)	Yes
Modern Countries	Controlled vocabulary, Multiple	Modern countries describing the provenance of the dataset	Yes
Network Topics	Text, Multiple	List of general network semantics, for example, face-to-face communication, communication at a distance	Yes
Node Topics	Text, Multiple	List of node semantics, for example, physical person, location, artefact	Yes
Edge Topics	Text, Multiple	List of edge semantics, for example, kinship, friendship, workplace collaboration, spatial closeness	Yes
Node Attributes	Text, Multiple	List of node attributes covered, for example, gender, occupation, geographic coordinates, source of information	No
Edge Attributes	Text, Multiple	List of edge attributes covered, for example, edge type, source of information, uncertain	No
Uncertain Nodes	Controlled vocabulary	Indication of whether uncertain nodes are present	No
Uncertain Node Attributes	Controlled vocabulary	Indication of whether uncertain node attributes are present	No
Uncertain Edges	Controlled vocabulary	Indication of whether uncertain edges are present	No
Uncertain Edge Attributes	Controlled vocabulary	Indication of whether uncertain edge attributes are present	No
Description of Uncertainty	Text	Description of uncertainty associated with the data	No

Note: Of these 13 attributes, six are mandatory to ensure basic semantic documentation, while seven are optional attributes that provide enhanced detail about node and edge characteristics and data uncertainty.

uncertainties that often accompany network data, particularly in historical or observational contexts.

Technical parameters

Most of the technical attributes are optional, since not all projects collect every technical detail; we strongly encourage contributors to publish quantitative descriptors whenever possible. Doing so enhances the analytical value of the dataset and supports more robust comparative and computational research. The technical attributes are divided into Network Structure and Network Size.

The *Network Structure* attributes concern the fundamental graph-theoretic properties that determine how data are organised and what analytical methods can be applied. They include basic characteristics like directionality and edge weights to more complex structural features, such as multilayer configurations and temporal dynamics. Of these 11 structural attributes, listed in Table 4, only two are mandatory (directionality and network weight), while nine optional attributes capture specialised network types, including multilayer, temporal, and spatial configurations.

The *Network Size* attributes capture the fundamental dimensions of each network (see Table 5). It includes two mandatory descriptors, Number of Nodes and Number of Edges, which provide essential information about the network's scale. Additionally, the optional Isolates attribute indicates how many nodes are not connected to any others. Although small in number, these attributes are technically significant and form the basis for understanding the overall structure and complexity of the network, and are reported in all network science publications and network data repositories.

Network construction and provenance

This section captures the details of how the network was constructed, providing essential context for understanding its provenance and the reliability of represented information (see Table 6). Contributors are expected to include high-level descriptions of the data curation process, outlining the actions taken during network construction and their implications, for example, "*only named entities which could be disambiguated against Wikidata were considered*." Each dataset must also include a list of sources used to create the network, along with detailed descriptions and source types. These may include bibliographic references or links to primary data repositories or publications that detail the source material. Where applicable, contributors should provide a link to the code used for data curation or to a publication that explains the curation methodology. An optional field for additional comments allows contributors to note caveats, elaborate on relational models or processes underlying the network, and suggest how the data might be reused.

Past Networks Repository

The *Past Networks Repository* is an online portal for discovering open-access datasets related to the human past, hosted by Aarhus University (see Figure 1). The repository does not store datasets

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

Table 4. The structural characteristics containing the structural properties of network datasets, specifying fundamental graph-theoretic features such as directionality, edge weighting, and topological complexity

Attribute name	Type	Description	Mandatory
Directionality	Controlled vocabulary	Do the network edges have a direction (directed), or not (undirected), or some edges do, and some do not (mixed)?	Yes
Weighted	Controlled vocabulary	Do the edges have numeric attributes? Example: a road network with intersections as nodes and road segments as edges, with segment length as an edge attribute	Yes
Multigraph	Controlled vocabulary	Does the network allow multiple edges between a pair of nodes? Example: a network of epistolary exchanges between individuals, with one edge per type of letter sent	No
Self-Loops	Controlled vocabulary	Does the network allow self-loops? A self-loop is an edge connecting a node to itself	No
Spatial	Controlled vocabulary	Spatial networks are networks where nodes and/or edges are explicitly defined based on their geographical features	No
Longitudinal	Controlled vocabulary	A longitudinal network includes information about the change of network structure through time	No
Multilayer	Controlled vocabulary	A multilayer network has multiple kinds of relations between its nodes. Example: a regional transportation network whose settlements are connected via roads, navigable rivers, etc.	No
Hypergraph	Controlled vocabulary	A hypergraph is a network whose edges may connect any number of nodes. Example: a network of political entities with edges representing alliances	No
Signed	Controlled vocabulary	A network is signed when the relation between two nodes can be attributed a positive or negative value. Example: a network of friends and enemies	No
Probabilistic	Controlled vocabulary	A network whose edges exist with a given probability	No
Multipartile	Controlled vocabulary	Does the network have nodes that are distinguishable in two or more categories, and connections cannot exist between nodes in the same category?	No

Note: Of these 11 structural attributes, only two are mandatory, while nine optional attributes capture specialised network types, including multilayer, temporal, and spatial configurations.

Table 5. Core attributes describing the quantitative dimensions of network datasets

Attribute name	Type	Description	Mandatory
Number of Nodes	Numeric	The number of nodes in the network	Yes
Number of Edges	Numeric	The number of edges in the network	Yes
Isolates	Numeric	The number of isolates in the network	No

Note: The number of nodes and edges are mandatory attributes, while the number of isolates is optional and provides additional insight into network connectivity.

Table 6. Attributes describing the network construction process, including mandatory documentation of sources and methodology

Attribute name	Type	Description	Mandatory
Description	Text	High-level description of the data curation process, including implications	Yes
Sources	Child entity, Multiple	Sources used to create the network; at least one source required	Yes
— *Source Description*	Text	The source(s) used to create the data; may include bibliography and links to data	Yes
— *Source Types*	Text, Multiple	The source types for the described source	Yes
Additional Comments	Text	Additional comments, caveats, or details about the data and guidance for reuse	No
Link to Curation Code	Hyperlink	Link to the code used for curation, if applicable	No

Note: Optional fields provide further context and support for reproducibility and reuse.

directly, but provides metadata and links to external storage locations and serves as a point of discovery for practitioners and students of network research of the past.

The core focus of the *Past Networks Repository* is on collecting information about networks of the human past, setting it apart from other network repositories and collections. Any data that can be used in past networks research for the study of social networks can be added, including historical and archaeological sources that either directly document interpersonal connections or from which such connections can be inferred or studied. Time, historical context, and data uncertainties are essential features of how data are structured, searched, and interpreted. To ensure chronological consistency and support for meaningful comparisons across datasets, contributors are required to specify both from and to

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

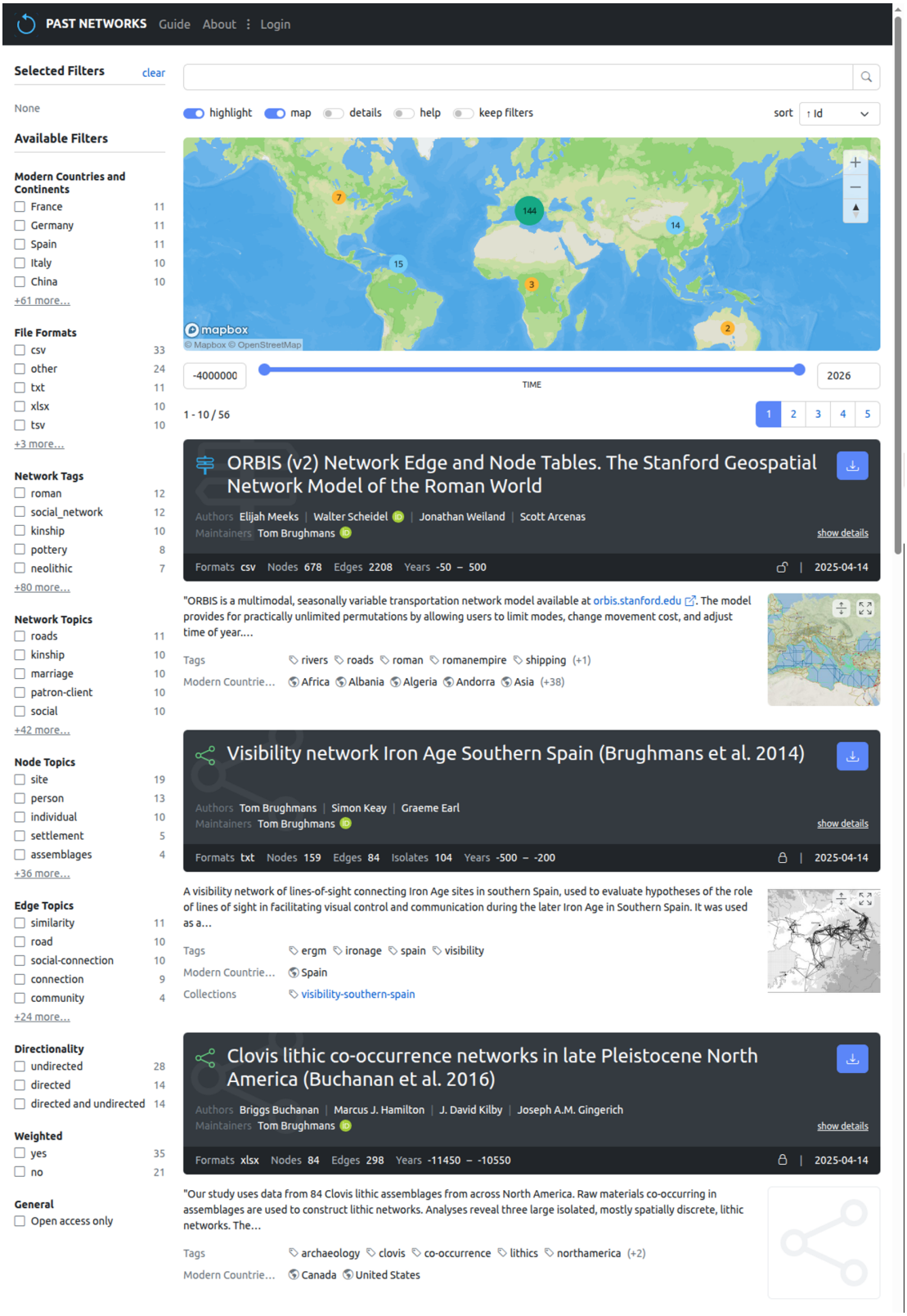


Figure 1. The *Past Networks Repository*landing page, providing an interactive search interface for the discovery of past network data. Screenshot taken on 5 March 2026.

years for each record. This design supports a wide range of time scales, accommodating projects that span from the distant past to more recent history. There is no concrete temporal focus to ensure that data relevant to the whole of the human past can be added – the data can represent networks from two million years ago, as well as from 2025. This flexibility makes the repository broadly applicable while maintaining a clear temporal structure and ease of data entry and search.

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

The repository is designed to be self-explanatory, requiring no manuals, and prioritises usability, standards compliance, and flexibility for non-standard data, including datasets with uncertain parameters. The main page, as seen in Figure 1, represents a primary point of access, where users can browse datasets through various visualisations, such as lists, maps, timelines, topic clusters, and co-authorship networks, and filter results interactively. The repository's design is grounded in the community-defined minimal documentation requirements, complemented by optional features that reflect best practices and aspirational standards. In cases where community recommendations were not followed, this was a deliberate choice to preserve simplicity and avoid unnecessary complexity, such as refraining from integrating third-party tools like Zotero, which would require a more dependent and intricate system architecture. However, with the current design, users still retain flexibility and can include DOIs, URLs, or formatted references using a free-form text field to preserve informational richness.

Technical infrastructure

The *Past Networks Repository* is built as a hybrid JSON and relational database using SQLite, designed for robustness and minimal future maintenance to ensure long-term sustainability. SQLite was specifically chosen for its archival stability; as a self-contained, serverless database, it ensures that research data remain highly portable and readable over the long term without the overhead of complex server administration.

The platform's overall architecture leverages *Node.js* for a fast, scalable backend and Svelte for a high-performance, reactive user interface.

The repository is hosted on Aarhus University servers and managed by the *Center for Humanities Computing* (CHC), Aarhus University. The infrastructure adheres to university policies for research data and information security, ensuring confidentiality, integrity, and accessibility for as long as there is funding.

Development and testing

The *Past Networks Repository* was developed through an iterative process involving *Past Social Network Project* team members during the alpha stage and volunteer practitioners during the beta stage, including a closed beta test in April–May 2025 by the former participants of the 2023 Aarhus workshop. This development approach focused on implementing community desirables, enabling data replication, supporting reuse, and ensuring clarity in both the documentation and the underlying data structure. Continuous feedback helped shape a user-friendly and reliable platform tailored to the needs of the past networks research community. Throughout its development, the repository was presented at various academic fora, including meetings, workshops, and conferences, where researchers and practitioners were invited to provide feedback. These engagements played a key role in refining the platform and aligning it with community expectations and scholarly practices.

Access and user management

Browsing the *Past Networks Repository* is free and open to all. To contribute data, users must have an active ORCID profile. Login and profile setup are handled via ORCID authentication, allowing users to securely access the platform using their existing credentials (see Figure 2). By logging in, users agree to abide by the site's Code of Conduct and agree to the Terms of Use. To enhance author discoverability and ensure data consistency, ORCID IDs are used not only for login but also as the primary method of identifying dataset creators and contributors, providing researchers with a unique, persistent digital identifier that disambiguates their work from others with similar names, ensures proper attribution across publications and academic activities, and improves the visibility of their contributions across publications and academic platforms.

For legacy data predating the widespread use of ORCID, identification by name and/or institutional affiliation is permitted. However, the use of ORCID is strongly encouraged as a best practice.

Data entry and data management

Once authenticated via ORCID, users can create and manage records (maintainer role), as well as invite other registered users to collaborate on their records. The maintainer of a record does not necessarily have to be one of the authors who created the original dataset or the publication. However, the maintainer will be responsible for the quality of the record entered into the repository. The maintainer can draft, publish, edit, and download the data they created as an ndJSON file, including mapping of Past Networks data structure to established Linked Data standards (PN JSON-LD). Maintainers can be informed of incorrect information on the records they maintain via their ORCID profile. Importantly, only the original maintainer, who created the record or a repository administrator, has the authority to delete it. The repository administrator has complete control over the data and user management and can delete them if they don't abide by the Code of Conduct. The project co-directors are the default repository administrators, and over time, the repository will transition to a community-driven administration and quality control model.

To facilitate projects involving multiple related datasets, the repository supports record duplication. This feature allows maintainers to replicate essential identification information from existing records, streamlining the data entry process. However, technical parameters must be entered manually to ensure each record remains unique and accurately reflects the specific network being documented. This approach enables users to efficiently record multiple networks within a single project while maintaining data integrity.

To prevent the creation of duplicate records, a warning system has been implemented: when creating a new record with a Data URL (a unique URL pointing to the location of the dataset) that already exists in the repository, an automated warning will be displayed alongside a link to the existing record. The user must then decide whether the records are duplicate or warrant a new entry.

Standardisation versus flexibility

The design of the *Past Networks Repository* carefully balances the need for standardised formats with the realities of historical datasets, which often contain uncertainties and reflect the topical diversity of the research community. Rather than enforcing rigid structures, the repository accommodates variation while maintaining usability and searchability.

One of the community-requested features that balances user freedom with input standardisation is the "suggester." Users are presented with a list of existing categories, or tags, drawn from the existing data, sorted from most to least popular, which they can choose from. If a suitable category does not exist, users can create new ones, which are then added to the list for future use.

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

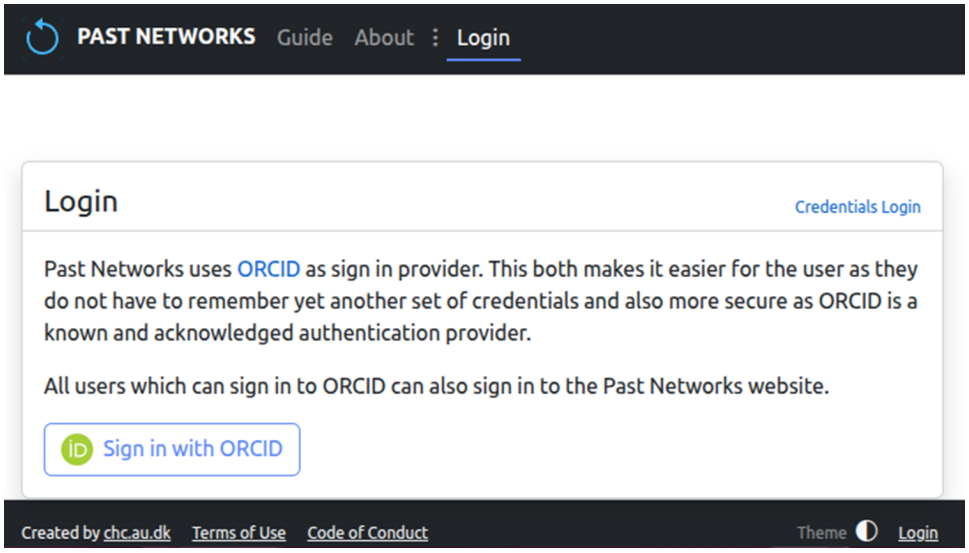


Figure 2. Login page of the *Past Networks Repository* using ORCID authentication. Screenshot taken on 5 March 2026.

This dynamic tagging system ensures consistency where possible, while allowing flexibility for unique or emerging topics.

The "suggester" functionality is used to populate project-level tags that define the scope of a dataset, as well as network-level topics; for details, see the "Network Topic" section. Users can also group related datasets by assigning them to the same project tag, enabling the creation of collections that reflect broader research initiatives or thematic clusters.

To further bridge project-specific flexibilities and established Digital Humanities standards, the repository aligns its internal metadata schema with widely adopted external vocabularies. Historical standardisation initiatives, such as Dublin Core (DCMI) and Schema.org, provide a stable infrastructure for semantic interoperability across domain-specific boundaries. The resulting machine-actionable schema files and their corresponding documentation are openly accessible via the web portal at https://pastnetworks.org/terms/ and can be downloaded directly with every dataset export. This semantic alignment ensures that flexible, user-generated categories remain fully discoverable and reusable within the broader linked data ecosystem.

Search and filtering

Discovery is central to the design of the *Past Networks Repository*, and users are provided with multiple intuitive ways to explore the data: browse, faceted search, full-text search with a built-in query language for advanced querying, map search, and chronological search. As seen in an example record in Figure 3, users can filter networks easily based on their properties, such as topics, types of networks, and keywords, using pre-populated filters and tag-based suggester fields. A time slider allows for temporal exploration of datasets, while a spatial map provides a visual overview of dataset locations. For more advanced use, the repository supports full-text search and Query Syntax, complete with eloquent examples to guide precise queries, available under the "help" toggle option below the search box at the repository's landing page. Additionally, results can be sorted by criteria such as ID, name, or creation date, enabling users to quickly locate relevant records. This multi-layered approach to data access ensures that both casual browsers and expert researchers can navigate the repository efficiently and uncover meaningful connections across historical networks.

Licensing and data export

Licensing and data rights are central to the mission of the *Past Networks Repository*. Therefore, we strongly encourage practitioners to publish their datasets and findings under open licenses, such as the Creative Commons licenses, as a standard practice. Open licensing promotes transparency, reuse, and collaboration across the research community. However, the *Past Networks Repository* also accommodates research data published under a restricted access option for cases where licensing constraints, privacy concerns, or institutional policies necessitate limited availability.

Metadata for individual dataset records contain their own licensing information. The responsibility for data quality, accuracy, and updates lies with the original record creators, not with the *Past Networks Repository*. The repository serves as a discovery platform and does not assume liability for errors or outdated information.

All information and metadata collected by the *Past Networks Repository* are available for export via the *Past Networks Repository* website (https://pastnetworks.org/about) under the *No Rights Reserved* Public domain license (CC0). Each night, a full export of all published network records is automatically generated in ndJSON format, with each line representing a single JSON network object. The export includes mapping of Past Networks data structure to established Linked Data standards (PN JSON-LD).

Desiderata: Community aspirations and future possibilities

Throughout the development of the *Past Networks Repository*, the community has expressed interest in a number of features that, while valuable, have not yet been implemented due to considerations of cost, sustainability, and technical complexity.

Dataset feedback mechanisms

A frequently requested feature is the ability for users to leave feedback or comments on datasets. While technically feasible, implementing such a system would require robust infrastructure, likely resembling a discussion forum restricted to registered users. Developers would also need to address potential security concerns, such as spam or compromised accounts, making this a non-trivial addition. Thus, it was ultimately decided that substantial

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

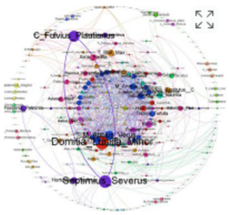
PAST NETWORKS Guide About ⋮ Login

Named individuals in stamped bricks in the Roman world

Authors Shawn Graham
Maintainers Petra Hermankova

Formats xlsx Nodes 0 Edges 0 Years 1 – 400 | 2025-05-19

The data used for network analysis in this context primarily comes from Roman stamped bricks, particularly those found in the Tiber Valley, which contain rich, multi-dimensional information such as the names of landowners (domini) and brick makers (officinatores), and details about the property or clay source (figlina) and accompanying figurative devices (signal). From these details, networks are constructed where specific pieces of information or entities become the nodes and their recorded relationships (like co-occurrence on a stamp or shared connections) become the links or edges.



Tags	brick-stamps bricks roman workshops
Modern Countries ...	Italy
Structure	Directionality directed Weighted no Hypergraph - Longitudinal - Multigraph - Multilayer - Multipartite - Probabilistic - Self Loops - Signed - Spatial -
Data Url	https://figshare.com/articles/dataset/CIL_XV_1_Instrumentum_Domesticum_a_working_copy_/92031?file=91252
Canonical Citation	Shawn Graham, On Connecting Stamps – Network Analysis and Epigraphy, Les nouvelles de l'archéologie, 135 | 2014, 39-44. https://doi.org/10.4000/nda.2353
License	Restricted Access
Data Publisher	Figshare
Related Publications	Shawn Graham, On Connecting Stamps – Network Analysis and Epigraphy, Les nouvelles de l'archéologie, 135 | 2014, 39-44.
Network Topics	economic epigraphy social
Node Topics	person
Edge Topics	material-co-occurrence
Node Attributes	-
Edge Attributes	-
Uncertainties	Nodes - Edges - Node Attributes - Edge Attributes -
Statistics	Avg. Clustering Coefficient - Avg. In Degree - Avg. Out Degree -

Construction

This dataset was created by manually digitising CIL volume XV.1 by the author as a graduate student. The data is a rough and necessarily incomplete transcription of the information contained in the Corpus Inscriptionum Latinarum XV.1 Instrumentum domesticum, edited by H. Dressel, 1891. It should not be used to identify stamped bricks (one should use instead Steinby's 'Indici complementari ai bolli doliari urbani (CIL XV.1) Acta Instituti Romani Finlandiae 11). Its utility lies in being able to be searched quickly, to narrow down the range of possibilities. I offer it in that spirit. It took author approximately 50 hours to produce.

Sources	Corpus Inscriptiom Latinarum. Specifically, the author's "working copy" of the CIL XV.1 'Instrumentum Domesticum' data, which was used for the analysis of relationships between figlinae and signa in brick stamps, is mentioned as being available online. Source Types inscriptions
Add. Comments	The article focuses on applying this method to explore two main types of networks: those connecting named individuals involved in the industry (like domini linked by shared officinatores) and those connecting stamp elements (figlinae linked by common signa), providing new ways to understand the organisation and logistics of brick production.

Figure 3. Metadata record for *Named individuals in stamped bricks in the Roman world* (Graham 2014) as displayed in the *Past Networks Repository* [record ID 19], maintainer of the record Petra Heřmánková. Screenshot taken on 5 March 2026.

feedback should be handled directly between users. Contributors are identifiable, and users are encouraged to contact them individually if they have meaningful input.

Data quality and curation

Ensuring the quality and reliability of datasets is a key concern for the *Past Networks Repository*. While a series of automated checks has been implemented to ensure that data are entered in the correct format, the validity and accuracy of the content itself remain the responsibility of the contributors. This approach balances usability with flexibility, allowing for a wide range of historical data while maintaining basic structural consistency.

Looking ahead, the repository aims to transition from administrator-led oversight to a more community-driven

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press

quality-control model. Initiatives such as the *Barcelona Past Networks Summer School* (BPNSS), or *The Connected Past* (TCP) and the *Historical Network Research* (HNR) communities may play a central role in this process. To support such a shift, enhanced user management features, such as account creation, password recovery, and role-based permissions, would be necessary.

To further support transparency and encourage high-quality contributions, the repository is considering implementing either a data completeness indicator or a tiered accreditation system. The completeness indicator could take the form of a visual progress bar, reflecting the proportion of fields populated. Alternatively, a tiered system, inspired by *Standards for Networking Ancient Prosopographies* (SNAP:DRGN) model or the Berners-Lee 5-Star LOD scheme (Berners-Lee 2006), could be adapted to suit the repository's needs. Over time, such a framework could help institutionalise standards, incentivise contributors, and support scholarly recognition. However, implementing this system would require a substantial time commitment from community members, both to define the ranking criteria and to apply them consistently across datasets.

Conclusion

This article addresses key challenges impeding large-scale research into past human social network structure, specifically the lack of standardised documentation and inconsistent community adoption of data practices. To resolve these issues, the authors propose past network data standards that align with current usage, identify missing components, and adhere to FAIR and LOUD data principles. Because past network research is inherently interdisciplinary, bridging disciplines such as history, archaeology, physics, and network science, these standards are vital to improve data provenance, licensing clarity, and analytical applicability across fields, while supporting reproducibility and enabling automated data reuse.

Crucially, the *Past Social Networks Project* developed the *Past Networks Repository* (https://pastnetworks.org/) to operationalise these standards, facilitate community adoption, enhance resource discoverability, and foster collaborative research dissemination. By setting up a sustainable and accessible infrastructure, this initiative directly confronts key challenges identified through community engagement: the necessity of a cultural shift towards meticulous documentation of dataset transformation and analysis processes, and the technical hurdles arising from conceptual variability and the absence of agreed-upon standards. The repository's design is meticulously shaped by community-defined minimal documentation requirements, mandating 27 out of 55 identified attributes for essential identification, authorship, network data properties, topics, and construction. The platform prioritises simplicity, intuitive use, and adaptability for historical data, balancing standardisation with flexibility. While automated checks ensure format consistency, the responsibility for data quality and accuracy rests with contributors, with future aspirations for a community-driven quality control model.

By championing transparency, reusability, and robust documentation, this framework provides essential infrastructure for those working with past network data and advances open science in the humanities. It enables impactful, collaborative, and interdisciplinary inquiries into the human past, facilitating systematic discovery and comparison while preserving methodological flexibility and academic rigour.

Data availability statement. All information and metadata collected by the *Past Networks Repository* are available for export via the *Past Networks Repository* website, under the *No Rights Reserved* Public domain license (CC0). Each night, a full export of all published network records is automatically generated in ndJSON format, with each line representing a single JSON network object, and mapping to established Linked Data standards (PN JSON-LD). Metadata for individual dataset records contain their own licensing information.

Acknowledgements. This article was inspired by a workshop held at Aarhus University (Denmark) in January 2023, organised by the *Past Social Networks Project*, and a similar workshop held at the Santa Fe Institute (United States) later that year. We would like to thank participants of both workshops for their feedback and valuable comments.

Author contributions. Conceptualisation: P.H., T.B., and M.M.; Data curation: All authors; Formal analysis: None; Funding acquisition: T.B.; Investigation: P.H., T.B., and M.M.; Methodology: P.H., T.B., and M.M.; Project administration: P.H. and T.B.; Resources: P.H., T.B., M.M., and P.B.V.; Software: P.B.V. and P.M.-D.; Supervision: P.H. and T.B.; Validation: P.H., T.B., and M.M.; Visualisation: P.B.V.; Writing – original draft: All authors. Writing – review and editing: All authors approved the final manuscript.

Funding statement. This work was made possible thanks to *The Carlsberg Foundation's Young Researcher Fellowship* (CF21-0382) awarded to the *Past Social Networks Project*.

Competing interests. The authors declare none.

Declaration of artificial intelligence use. AI-assisted tools, such as Grammarly, Overleaf AI, and Claude Sonnet 4.6, were used for proofreading purposes. All intellectual content, original ideas, their execution, and any errors remain the sole responsibility of the authors.

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Bekiari, Chryssoula, George Bruseker, Erin Canning, et al. eds. 2024. "Definition of the CIDOC Conceptual Reference Model." Version 7.1.3. CIDOC CRM Special Interest Group. Accessed August 26, 2026. https://cidoc-crm.org/get-last-official-release.

Berners-Lee, Tim. 2006. "Linked Data–Design Issues." https://www.w3.org/DesignIssues/LinkedData.html.

Bodard, Gabriel, Hugh Cayless, Chiara Cenati, et al. 2021. "Modeling Epigraphy with an Ontology." 10.5281/ZENODO.4639507.

Brughmans, Tom, Petra Heřmánková, Matteo Mazzamurro, et al. 2025. "Past Networks: The Open Online Repository Portal for Network Data About the Human Past." https://pastnetworks.org.

Brughmans, Tom, Barbara J. Mills, Jessica Munson, et al. eds. 2023. *The Oxford Handbook of Archaeological Network Research*, Oxford Handbooks. Oxford University Press. 10.1093/oxfordhb/9780198854265.001.0001.

Brughmans, Tom, and Matthew A. Peeples. 2023. *Network Science in Archaeology*, Cambridge Manuals in Archaeology. Cambridge University Press. 10.1017/9781009170659.

Burrows, Toby. 2023. "Reproducibility, Verifiability, and Computational Historical Research." *International Journal of Digital Humanities* 5, no. 2: 283–98. 10.1007/s42803-023-00068-9.

CIDOC CRM Special Interest Group. 2006. "CIDOC Conceptual Reference Model." Accessed August 26, 2026. https://cidoc-crm.org/.

Clauset, Aaron, Ellen Tucker, and Matthias Sainz. 2016. "The Colorado Index of Complex Networks." https://icon.colorado.edu.

Collar, Anna, ed. 2022. *Networks and the Spread of Ideas in the Past: Strong Ties, Innovation and Knowledge Exchange*, Digital Research in the Arts and Humanities. Routledge. 10.4324/9780429429217.

Davis, Timothy A., and Yifan Hu. 2011. "The University of Florida Sparse Matrix Collection." *ACM Transactions on Mathematical Software* 38, no. 1: 1–25. 10.1145/2049662.2049663.

Deckers, Pieterjan, Eero Hyvönen, Michael Lewis, et al. 2026. "ARCH-ON: A New Ontological Framework to Describe Archaeological Objects for Digital Humanities Research." *Digital Humanities in the Nordic and Baltic Countries Publications* 8, no. 1. 10.5617/dhnbpub.13208.

Dublin Core Metadata Initiative. 1995. "Dublin Core Metadata Initiative." Accessed August 26, 2026. https://www.dublincore.org/.

Elliott, Tom, Gabriel Bodard, Hugh Cayless, et al. 2006–2026. "EpiDoc: Epigraphic Documents in TEI XML." Accessed August 26, 2026. https://epidoc.stoa.org/.

Graham, Shawn. 2014. "On Connecting Stamps—Network Analysis and Epigraphy." *Les Nouvelles de l'archéologie*, no. 135: 39–44. 10.4000/nda.2353.

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press. 10.1142/p981.

Heřmánková, Petra, Imran Asif, Marietta Horster, et al. 2025. "From Fragmented Data to Linked History: Developing the FAIR Epigraphic Vocabularies." *Journal of Open Humanities Data* 11, no. 1. 10.5334/johd.428.

Huskey, Samuel J. 2024. "Committing to Reproducibility and Explainability: Using Git as a Research Journal." *International Journal of Digital Humanities* 6, no. 1: 9–21. 10.1007/s42803-023-00076-9.

Hyvönen, Eero. 2023. "Digital Humanities on the Semantic Web: Sampo Model and Portal Series." *Semantic Web* 14, no. 4: 729–44. 10.3233/SW-223034.

Joyeux-Prunel, Béatrice. 2024. "Digital Humanities in the Era of Digital Reproducibility: Towards a Fairest and Post-computational Framework." *International Journal of Digital Humanities* 6, no. 1: 23–43. 10.1007/s42803-023-00079-6.

Laguna-Palma, David, and Pablo Barruezo-Vaquero. 2024. "Advancing the LOUD+FAIR Data Principles in Eastern Mediterranean Archaeological Research: Insights from the PERAIA Project." *Advances in Archaeological Practice* 12, no. 4: 359–74. 10.1017/aap.2024.16.

Leskovec, Jure, and Rok Sosič. 2016. "SNAP: A General-Purpose Network Analysis and Graph-Mining Library." *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, no. 1: 1. 10.1145/28983.

Library of Congress. 1999. "MARC 21 Format for Bibliographic Data." Accessed August 26, 2026. https://www.loc.gov/marc/bibliographic/.

Marwick, Ben. 2017. "Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation." *Journal of Archaeological Method and Theory* 24, no. 2: 424–50. 10.1007/s10816-015-9272-9.

Marwick, Ben, Jade D' Alpoim Guedes, C. Michael Barton, et al. 2017. "Open Science in Archaeology." *SAA Archaeological Record* 17: 8–14. 10.17605/OSF.IO/3D6XX.

Mazzamurro, Matteo, Tom Brughmans, Albert Diaz-Guilera, et al. 2026. "Network Science Education for the Humanities: The Barcelona Past Networks Summer School." In *2026 IEEE Global Engineering Education Conference (EDUCON)*, 1–10. IEEE. 10.1109/EDUCON67543.2026.11574492.

Middle, Sarah. 2023. "A Documentation Checklist for (Linked) Humanities Data." *International Journal of Digital Humanities* 5, nos. 2–3: 353–71. 10.1007/s42803-023-00072-z.

Murano, Francesca, Achille Felicetti, and Pavlos Fafalios. 2023. "Definition of the CRMtex: An Extension of CIDOC CRM to Model Textual Entities." Accessed August 26, 2026. https://cidoc-crm.org/sites/default/files/CRMtex_v2.0_June_2023.pdf.

Nicholson, Christopher, Sarah Kansa, Neha Gupta, et al. 2023. "Will It Ever Be FAIR?: Making Archaeological Data Findable, Accessible, Interoperable, and Reusable." *Advances in Archaeological Practice* 11, no. 1: 63–75. 10.1017/aap.2022.40.

Olsson, Michael, Olle Sköld, and Lisa Andersson. 2024. "Layers upon Layers: Data Sharing & Reuse Challenges in Archaeological Contexts." *Information Research an International Electronic Journal* 29, no. 2: 454–63. 10.47989/ir292840.

Prag, Jonathan, and Imran Asif. 2026. "FAIR Epigraphic Ontology (EPONT)." https://ontology.inscriptiones.org/epont/.

Sobotkova, Adela, and Petra Hermankova. 2022. "Emergent Archaeological Realities and Reusable Datasets: What is the Priority?" In *Digital Heritage and Archaeology in Practice*, edited by **Ethan Watrall and Lynn Goldstein**, 135–49. Florida University Press. 10.5744/florida/9780813069302.003.0007.

TEI Consortium. 1987. "Text Encoding Initiative." Accessed August 26, 2026. https://tei-c.org/.

TEI Consortium. 2026. "TEI P5: Guidelines for Electronic Text Encoding and Interchange." Version 4.12.0. TEI Consortium. Accessed August 26, 2026. https://www.tei-c.org/Guidelines/P5/.

Thiery, Florian. 2019. "Sphere 7 Data: LOUD and FAIR Data for the Research Community." 10.5281/ZENODO.2643470.

Weiß, Lea, Laura Von Welczeck, Malte Vogl, et al. 2025. "Past, Present, and Future of HNR: Reflections on the Practices and Methods in Historical Network Research Based on a Quantitative Survey." *Journal of Historical Network Research* 11, no. 1. 10.25517/JHNR.V11I1.103.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, no. 1: 160018. 10.1038/sdata.2016.18.

https://doi.org/10.1017/chr.2026.10051 Published online by Cambridge University Press