



A National Relational Dataset of Reconstructed Prehistoric and Ancient Buildings in Japan

JOHN ERTL 

GARY ROSS 

**Author affiliations can be found in the back matter of this article*

DATA PAPERS

]u[ubiquity press

ABSTRACT

This dataset documents modern architectural reconstructions of prehistoric and ancient buildings across Japan that are based primarily on archaeological remains. It contains records for 1,014 reconstructed buildings at 346 sites, including building type, materials, construction and rebuilding history; site names, locations, and heritage designations; designer attribution and institutional affiliation; and documentary sources. The data were assembled through documentary research, correspondence with site authorities, and field survey. The relational structure records reconstruction location separately from the archaeological evidence on which it is based and preserves uncertainty in key information. These data support the comparative study of reconstruction practices across time and place and of relationships among archaeological evidence, reconstructed architecture, institutions, and designers.

CORRESPONDING AUTHOR:

John Ertl

Faculty of Economics, Keio University, Tokyo, Japan

jjertl@keio.jp

KEYWORDS:

architectural reconstruction;
prehistoric architecture;
heritage sites; relational
database; Japan

TO CITE THIS ARTICLE:

Ertl J, Ross G 2026 A National Relational Dataset of Reconstructed Prehistoric and Ancient Buildings in Japan. *Journal of Open Archaeology Data*, 14: 24, pp. 1–8. DOI: <https://doi.org/10.5334/joad.232>

(1) OVERVIEW

CONTEXT

Reconstructed prehistoric and ancient buildings are prominent features of archaeological sites, museums, and historical parks throughout Japan, often providing visitors with their most immediate image of past settlements. These include pit dwellings, raised-floor storehouses, and other forms of prehistoric and ancient architecture reconstructed from archaeological evidence. Their form may also reflect comparative evidence, architectural interpretation, institutional decisions, and practical constraints. Studied comparatively, they offer a way to identify broader trends in archaeological practice and heritage reconstruction.

Excavation reports in Japan document buried archaeological features extensively, but the reconstructions produced from them are recorded far less consistently. Information about reconstructed buildings is scattered across archaeological reports, institutional publications, newspapers, official webpages, and other sources, and the amount of documentation varies greatly from one reconstruction to another. Some have been extensively recorded and discussed, while others are difficult to trace or survive only fragmentarily in the documentary record. What survives in that record therefore shapes which reconstructions come to appear representative of reconstruction practice as a whole.

The dataset consolidates a decade of survey work on reconstructed prehistoric and ancient buildings in Japan. The project began in 2016 with a survey of Jomon-period reconstructions [1], followed by a national survey that identified 984 reconstructed buildings at 340 sites [2]. A series of regional surveys subsequently expanded and updated the photographic and bibliographic record,

adding newly identified buildings and revisiting previously documented sites [3–8].

This project is the first sustained national survey effort since the 1978 survey of reconstructed ancient houses [9] to document reconstructed prehistoric and ancient buildings systematically across Japan. The present release organizes the resulting material as a building-level relational dataset linking reconstructed buildings with their sites, archaeological basis, designers, and documentary sources while preserving uncertainty in key information. To our knowledge, it is the first openly released relational dataset of this kind for reconstructed prehistoric and ancient buildings in Japan.

Reconstructions can stand in different relationships to the archaeological remains on which they are based. Some are built directly above the excavated features they represent, with the archaeological remains protected through vertical displacement. Others reconstruct a particular excavated building but are erected away from the original feature, either elsewhere at the same site or at another location. Model reconstructions, by contrast, draw on varying degrees of archaeological and comparative evidence without attempting to reproduce a particular excavated building. Recording these differences makes it possible to compare systematically how archaeological evidence relates to reconstructed architecture.

SPATIAL COVERAGE

Description: Japan; 346 reconstructed-building sites across 46 of 47 prefectures (Figure 2). Coordinates use WGS 84 (EPSG:4326).

Northern boundary: 45.1252405

Southern boundary: 26.391525

Eastern boundary: 145.116558

Western boundary: 127.994515

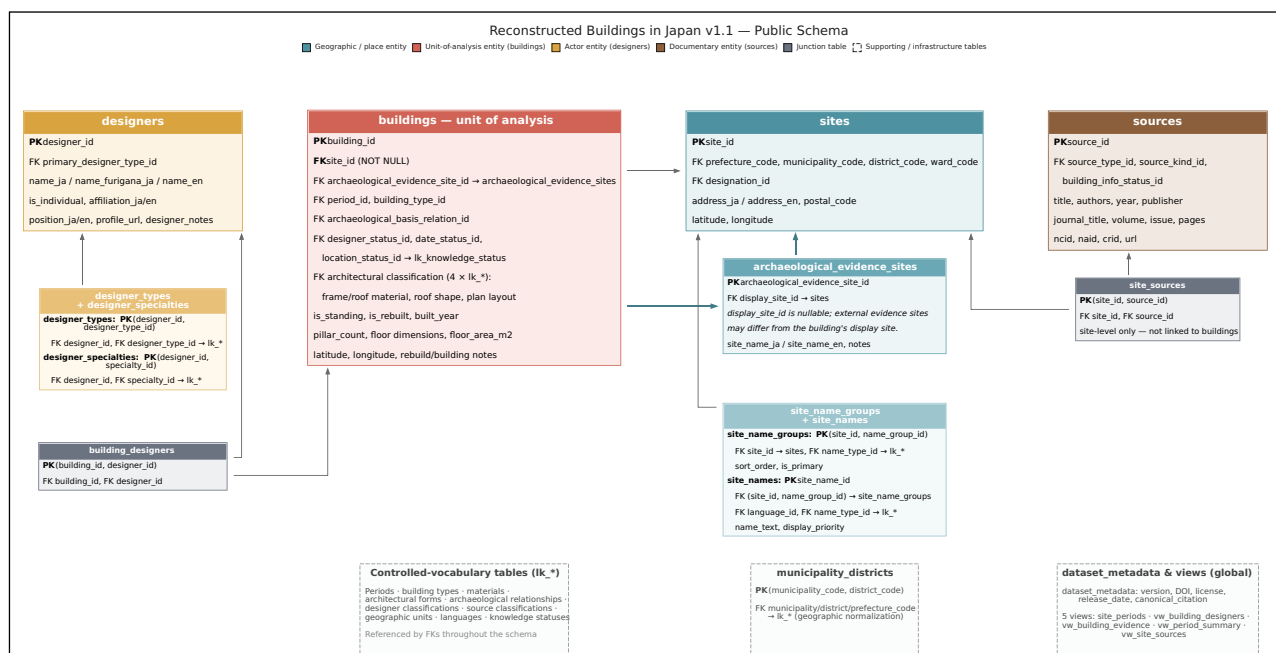


Figure 1 Entity-relationship diagram of the public dataset. Principal entities are reconstructed buildings, sites, designers, and documentary sources, with supporting tables for archaeological evidence, names, geography, and controlled vocabularies.

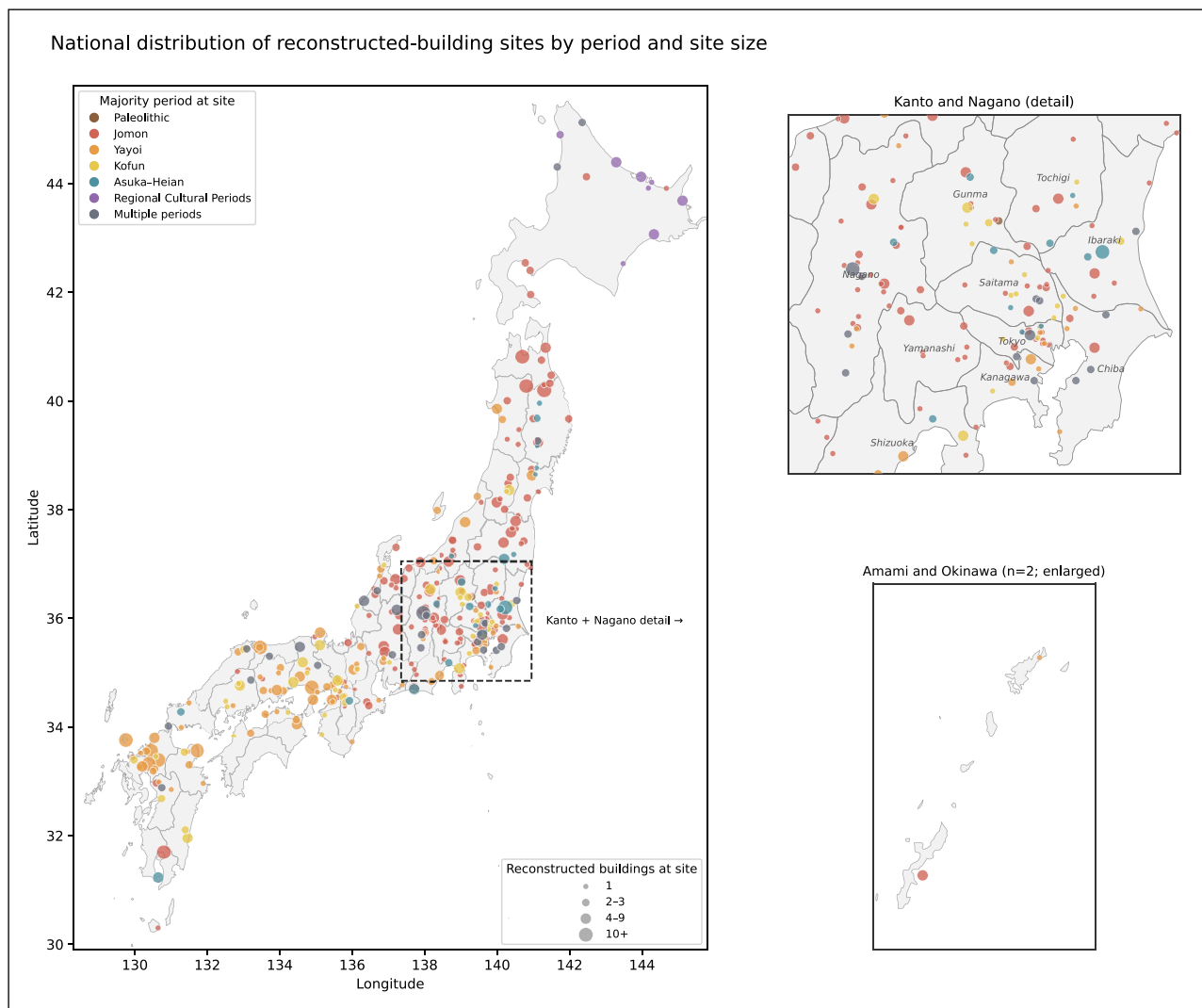


Figure 2 National distribution of reconstructed-building sites by archaeological period and site size. Symbol colour indicates the majority archaeological period group at each site, and symbol size indicates the number of reconstructed buildings. Insets show Kanto and Nagano, and Amami and Okinawa.

TEMPORAL COVERAGE

Represented archaeological periods: Paleolithic, Jomon (c. 14,000–900 BCE), Yayoi (c. 900 BCE–300 CE), Kofun (c. 250–600), Asuka (c. 600–710), Nara (710–794), and Heian (794–1185), together with region-specific chronologies such as Epi-Jomon (c. 300 BCE–700 CE), Okhotsk (c. 5th–9th centuries CE), and Satsumon (c. 7th–12th centuries CE).

Recorded construction dates: 1932–2023.

(2) METHODS

STEPS

Data collection focused on documenting reconstructed buildings and their characteristics, the sites where they are located, their archaeological basis, attributed designers, and the documentary sources used to establish this information. Research began by reviewing the 1978 national survey of reconstructed ancient houses [9] and was expanded through documentary

research, correspondence with site authorities, and visits to archaeological sites and museums.

Documentary research drew on excavation and site-development reports, academic and municipal publications, museum materials, newspapers, official webpages, and other available sources. Because information about reconstructed buildings was often incomplete or difficult to locate, over 150 institutions, including boards of education and other site authorities, were contacted to identify additional documentation and to establish or clarify information that could not be confirmed from published sources, including dates of construction, designer names, and publication records.

Visits to archaeological sites and museums were used to verify the presence, location, condition, and architectural characteristics of reconstructed buildings and to supplement gaps in the documentary record. Field photography was conducted at 304 of the 346 sites in Version 1.1 (88%), providing a visual record of the reconstructions at the time of observation. The photographic archive and its metadata are not included in this release.

Collected information was organized relationally around reconstructed buildings, the sites where they are located, archaeological evidence sites, designers, and documentary sources. This structure allows a site to contain multiple reconstructions, a designer to be associated with buildings at different sites, and a documentary source to be associated with one or more sites. Controlled vocabularies were used to standardize recurring categories such as archaeological period, building type, and materials, while information that could not be established was retained as unknown or unresolved rather than inferred.

The public release [10] was prepared from a larger working research database containing additional correspondence, research notes, administrative records, and photographic material that are not part of the public dataset. Relevant research data were selected and organized within the public relational schema, with SQLite as the canonical release format and CSV exports provided for accessibility. Scripts documenting creation of the public schema, population of its tables and views, and validation of the released database are included with the dataset. Because the private working database is not deposited, the public database cannot be regenerated independently from the released package alone.

SAMPLING STRATEGY

The dataset includes modern architectural reconstructions based primarily on archaeological evidence. Both complete and partial reconstructions are included, including structures in which only the architectural frame has been reconstructed. Reconstructions that are no longer standing are also retained.

Of the 1,014 reconstructed buildings in Version 1.1, 203 are recorded as no longer standing and 811 as standing. At the site level, 86 of the 346 sites have no standing reconstruction remaining, while 125 sites include at least one building recorded as no longer standing.

The project aims for comprehensive national coverage rather than statistical sampling. Because information about reconstructed buildings is widely dispersed and unevenly documented, no authoritative national register exists against which completeness can be independently demonstrated.

Reconstructions that are based primarily on historical or textual documentation rather than archaeological remains are excluded, including castles, palaces, temples, and administrative buildings. Preserved original buildings, virtual or graphic reconstructions, and ground-level outlines or markers that do not reconstruct architectural form are also excluded.

Reconstructions are classified by their relationship to the archaeological evidence on which they are based: Same Site, where the evidence comes from the site where the reconstruction is located; External Site, where it comes from another archaeological site; Model Reconstruction,

where the basis is comparative or generalized rather than a specific excavated building; and Unknown, where the relationship could not be established.

Rebuilding or replacement is recorded within the existing building record when it represents a continuation of the same reconstruction, while a new reconstruction based on a different building or evidentiary basis receives a new record.

QUALITY CONTROL

In most cases, no single source provided a complete record of a reconstructed building. Records were therefore assembled from multiple documentary sources, with information cross-checked wherever possible and site authorities consulted when further clarification was needed. Site visits provided an additional source of evidence, allowing documented information to be checked and observable characteristics to be recorded directly.

To avoid treating missing information as equivalent, the database requires a knowledge status for construction date, designer attribution, and location. Information is recorded as confirmed when supported by the available evidence, as confirmed unknown when investigation has established that the information cannot presently be recovered, and as under investigation when the issue remains unresolved. This allows missing data to be interpreted according to whether information is known to be unavailable or remains unresolved. In Version 1.1, construction date is recorded as confirmed for 984 buildings, confirmed unknown for 2, and under investigation for 28; designer attribution for 884, 39, and 91 respectively; and location for 979, 28, and 7.

Primary and foreign keys, field constraints, and controlled vocabularies were used throughout the relational schema to maintain consistent links between buildings, sites, designers, archaeological evidence sites, sources, and coded categories. Additional checks addressed problems that could not be prevented structurally, including duplicate records, inconsistent classifications, implausible construction dates and coordinates, and anomalous measurements.

The released SQLite database was subjected to final structural and referential-integrity checks. It passes SQLite's PRAGMA integrity_check and PRAGMA foreign_key_check with no violations, and the CSV exports were checked against the corresponding SQLite tables for consistency. The supplied validation script also tests key structural constraints and semantic relationships in the released database.

CONSTRAINTS

Documentation varies substantially across reconstructed buildings and sites, and some attributes are much more completely recorded than others. Detailed architectural variables such as dimensions, floor area, pillar count, and materials, as well as rebuilding history, were

not systematically collected as part of the original survey design but were recorded when documented information was available. Their coverage is therefore uneven. For example, frame material is recorded for all 1,014 buildings, roof material for 983, floor area for 367, and pillar count for 327. Analyses using incompletely documented variables should report the relevant denominator and should not assume that available records constitute a statistically representative subsample of all reconstructions.

The absence of a recorded source should not be taken as evidence that no documentation exists, only that none was identified through the project's methods. The absence of a recorded rebuilding event should not be interpreted as evidence that rebuilding did not occur.

Documentary sources are linked at the site level and distinguished according to whether they contain information on reconstructed buildings or only on the site and its archaeology. Sources are not linked to individual buildings or specific building characteristics because the documentation was generally collected and used at the site level.

Coordinates were identified primarily using Google Maps and are intended to provide approximate locations for mapping and comparative analysis. The number of decimal places stored should not be interpreted as indicating survey-grade precision. Because map imagery and alignment can change over time, stored coordinates may no longer correspond exactly to the apparent position of a building on current imagery.

Reconstructed buildings continue to change over time. Standing status, rebuilding history, and related information reflect the evidence incorporated through 9 August 2026, but per-record observation dates are not recorded in Version 1.1.

(3) DATASET DESCRIPTION

The public database is organized around four principal data tables: Sites, Buildings, Designers, and Sources (Figure 1). Version 1.1 contains 346 sites, 1,014 reconstructed buildings, 279 designers, and 708 source records.

Sites record the places associated with reconstructed buildings, together with address, coordinates, heritage designation, and standardized prefecture and municipality codes that support geographic searching and comparison. The primary name is usually that of the archaeological site, but the same place may also be identified through an associated park, museum, school, or other name. These alternatives are stored separately so that multiple valid forms can be retained without creating duplicate site records. Japanese and English forms are recorded together with phonetic readings (katakana) where available, and one name group is designated as primary for display and reference.

Buildings record individual reconstructed structures and their architectural and historical characteristics. These include building type, archaeological period, construction date, standing status, rebuilding history, dimensions, plan form, pillar count, and materials used for the frame, roof, and other architectural elements. Each building is linked to the site where it is located and classified according to its relationship to the archaeological evidence on which the reconstruction is based.

Designers record the individuals and organizations attributed to reconstructed buildings. Designers are stored as separate entities so that the same individual or organization can be linked to multiple buildings, while a building may also be associated with more than one designer. Designer types and institutional or professional affiliations are recorded where available, allowing reconstructions to be searched and compared by designer category as well as by individual name.

Sources record publications and other documentary materials associated with sites and their reconstructed buildings. These include excavation and site-development reports, academic and municipal publications, newspaper articles, official webpages, and other relevant materials. Sources that directly document reconstructed buildings are distinguished from those that provide broader archaeological or historical information about the site. URLs and bibliographic identifiers are recorded where available.

The database also records the archaeological provenance of reconstructed buildings by identifying the archaeological site and individual feature on which a reconstruction is based. Because these evidence sites are not always the same as the sites where reconstructions are located, they are stored in a separate archaeological evidence-sites table. Version 1.1 contains 397 archaeological evidence-site records, including the 346 reconstruction sites and 51 additional archaeological sites; 88 buildings are linked to a specific external evidence site.

Across the 1,014 buildings, the archaeological-basis classification records 553 as Same Site, 98 as External Site, 359 as Model Reconstruction, and 4 as Unknown. This allows the evidentiary relationship between excavated remains and reconstructed buildings to be traced across sites.

Controlled vocabularies standardize recurring categories across the database, including archaeological period, building type, materials, architectural form, heritage designation, archaeological basis, source and designer classifications, language, and knowledge status. The vocabulary tables and their definitions are included in the public release so that coded values can be interpreted consistently.

The SQLite database preserves the full relational structure of the dataset, including links, constraints, controlled vocabularies, and supporting tables. CSV exports are provided with stable identifiers allowing

records to be linked across files. The release also includes convenience views for commonly used relationships, together with documentation, validation scripts, and example queries.

OBJECT NAME

reconstructed_buildings_japan_v1.1.sqlite

DATA TYPE

Processed data; secondary data

FORMAT NAMES AND VERSIONS

SQLite 3; CSV (UTF-8 with BOM); Markdown; SVG; PNG; Python scripts; SQL scripts; CITATION.cff

CREATION DATES

26/07/2016–09/08/2026

DATASET CREATORS

John Ertl, Faculty of Economics, Keio University, ORCID [0009-0002-3581-1033](https://orcid.org/0009-0002-3581-1033) — dataset conception, data collection, curation, and project administration.

Gary Ross, Faculty of Pharmaceutical Sciences, Kanazawa University, ORCID [0009-0002-7878-133X](https://orcid.org/0009-0002-7878-133X) — database design, schema development, data curation, and validation.

LANGUAGE

Japanese and English

LICENSE

CC BY 4.0

REPOSITORY LOCATION

Zenodo. DOI: [10.5281/zenodo.23166842](https://doi.org/10.5281/zenodo.23166842)

PUBLICATION DATE

06/10/2026

(4) REUSE POTENTIAL

The dataset supports comparison of reconstruction practice across time and place by linking reconstructed buildings with construction dates, archaeological periods, architectural characteristics, heritage designation, and geographic location. Researchers can use these relationships to identify regional concentrations, compare changes in reconstruction practice over time, and examine whether particular architectural forms or materials are associated with specific periods or institutional settings.

The same structure is suitable for teaching. Students can begin with descriptive comparisons, such as mapping the regional density of reconstructions by

period or tracing one designer's work across multiple sites, and progress to questions requiring joins across tables, such as comparing reconstructions built above excavated features with those based on comparative evidence. The knowledge-status fields also provide an example of structured missingness for teaching data handling.

The distinction between reconstruction location and archaeological basis supports a second set of analyses. Because buildings are classified as Same Site, External Site, Model Reconstruction, or Unknown, researchers can compare architectural characteristics with the kind of archaeological evidence on which a reconstruction was based. This makes it possible, for example, to ask whether reconstructions based on a specific excavated building differ systematically from those based on comparative or generalized evidence, or whether these relationships change across periods or regions.

Designer and source records add further relational dimensions. Researchers can compare reconstruction practices by designer type, institutional affiliation, professional specialty, source type, and knowledge status, and can identify individuals or organizations involved in multiple projects. Source and knowledge-status data can also be used to examine where documentation is concentrated or incomplete rather than treating all missing information as equivalent.

The use of stable identifiers, normalized Japanese and English names, and separate controlled-vocabulary tables also provides a basis for future alignment with external authority resources. Designer records could, for example, be reconciled with name authorities such as the Library of Congress Name Authority File or the Getty Union List of Artist Names, while place records and architectural vocabulary could potentially be mapped to the Getty Thesaurus of Geographic Names and Art & Architecture Thesaurus. These external mappings are not included in Version 1.1, but the relational structure permits such identifiers or crosswalks to be added in future releases without altering the existing internal identifiers.

These uses are primarily quantitative and comparative. The dataset is designed to identify and describe patterns and trends in the distribution, form, archaeological basis, documentation, and institutional organization of reconstructed buildings. Explaining why those patterns developed requires qualitative investigation using archaeological, documentary, historical, and other contextual evidence.

DATA ACCESSIBILITY STATEMENT

Version 1.1 of the dataset is openly available on Zenodo under a CC BY 4.0 license at DOI: [10.5281/zenodo.23166842](https://doi.org/10.5281/zenodo.23166842).

ETHICS AND CONSENT

Ethics approval was not required for this research. The released dataset concerns reconstructed buildings, archaeological sites, designers, and documentary sources. Designer names and professional affiliations are recorded in their professional capacity from published or institutional sources. Information obtained through correspondence with site authorities was used to establish or verify factual records, but the correspondence itself and personal contact information are not included in the dataset.

ACKNOWLEDGEMENTS

The authors thank Yasuyuki Yoshida and Corey Noxon for their collaboration on the regional survey series that contributed to the development of this project. We also thank the staff of archaeological sites, museums, boards of education, and other heritage institutions across Japan who provided information, documentation, and assistance during the survey.

FUNDING STATEMENT

The research and data collection underlying this dataset were supported by JSPS KAKENHI Grant Numbers 15KK0043 (*Diversity in Archaeology and Cultural Resource Production in Japan*), 16K03219 (*Ethnography of Archaeological Production, Consumption, and Transformation*), 19H01394 (*Ethnography of Archaeological Excavation, Laboratory Analysis, and Site Development*), and 23K25438 (*Ethnography of Archaeology and Architectural Reconstruction*).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

John Ertl: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Visualization; Writing – original draft; Writing – review & editing.

Gary Ross: Data curation; Methodology; Software; Validation; Writing – review & editing.

AUTHOR AFFILIATIONS

John Ertl  orcid.org/0009-0002-3581-1033
Faculty of Economics, Keio University, Tokyo, Japan

Gary Ross  orcid.org/0009-0002-7878-133X
Faculty of Pharmaceutical Sciences, Kanazawa University,
Kanazawa, Ishikawa, Japan

REFERENCES

1. **Ertl J.** 縄文時代の復元建物の実態調査 [Survey of Jomon-period reconstructed buildings]. In: 一戸町教育委員会 [Ichinohe Town Board of Education] (ed.) 『御所野遺跡環境整備事業報告書 III—総括報告書』 [Goshono Site Environmental Development Project Report III: Comprehensive Report]. 一戸町 [Ichinohe]: 一戸町教育委員会 [Ichinohe Town Board of Education], 2017; pp. 67–79.
2. **Ertl J.** Survey of Prehistoric and Ancient Period Architectural Reconstructions in Japan. *Japanese Journal of Archaeology*. 2021;8:157–199.
3. **Ertl J, Yoshida Y.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in Kansai, Chugoku, and Shikoku Regions in Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2022;32:1–50.
4. **Ertl J, Yoshida Y.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in Hokuriku and Tokai Regions in Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2023;33:21–62.
5. **Ertl J, Yoshida Y, Noxon C.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in the Koshin-etsu Region of Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2024;34:1–56.
6. **Ertl J, Yoshida Y.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in Hokkaido and Tohoku Regions of Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2025;35:1–60.
7. **Ertl J, Yoshida Y.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in the Kyushu Region of Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2026;36:1–57.
8. **Ertl J, Yoshida Y.** Survey of the Present Conditions of Prehistoric Architectural Reconstructions in the Kanto Region of Japan. *The Keio University Hiyoshi Review of Social Sciences*. 2027;37. Forthcoming.
9. 奈良国立文化財研究所埋蔵文化財センター [Nara National Research Institute for Cultural Properties, Center for Archaeological Operations] 1978 古代家屋復元実態調査 [Survey of the Actual Conditions of Reconstructed Ancient Houses]. 埋蔵文化財ニュース [Maizō Bunkazai News / CAO News] 13. 奈良 [Nara]: 奈良国立文化財研究所埋蔵文化財センター [Nara National Research Institute for Cultural Properties, Center for Archaeological Operations].
10. **Ertl J, Ross G.** A National Relational Dataset of Reconstructed Prehistoric and Ancient Buildings in Japan, Version 1.1 [Dataset]. *Zenodo*; 2026. DOI: <https://doi.org/10.5281/zenodo.23166842>

TO CITE THIS ARTICLE:

Ertl J, Ross G 2026 A National Relational Dataset of Reconstructed Prehistoric and Ancient Buildings in Japan. *Journal of Open Archaeology Data*, 14: 24, pp. 1–8. DOI: <https://doi.org/10.5334/joad.232>

Submitted: 18 August 2026 **Accepted:** 07 October 2026 **Published:** 23 October 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Journal of Open Archaeology Data is a peer-reviewed open access journal published by Ubiquity Press.

