

Structured Data for Historians: Glossary of Terms

Plain English definitions for the terms used in structured data research. For use alongside the Wide Format and Multi-Line Format templates.

Term	Plain English Definition	Example
Claim	A statement made by a source. Not yet confirmed as fact.	The census says age 42. That is a claim. Whether it is accurate is a separate question.
Conclusion	A decision you make after weighing the evidence. A conclusion should always be recorded separately from the claim that led to it.	You conclude that John Smith in the 1851 census is the same man as John Smith in the 1841 census, based on age, address, and occupation.
Confirmed	An identity or fact that has been verified against at least one other independent source.	Identity confirmed — cross-referenced with baptism register 1823.
CSV	Comma-Separated Values. A plain text file where each column is separated by a comma. Opens in Excel, LibreOffice, or any spreadsheet software.	Source_Ref,Source_Type,Event_Date becomes three columns when opened.
Data label	The name you give to a column or field in your structured data. Should be clear, consistent, and use underscores instead of spaces.	Surname_Recorded_As not 'Surname as written'
Evidence	The source material you are working from. Evidence is what you have. A conclusion is what you draw from it.	The will is evidence. Your interpretation of the will is a conclusion.
Event	A dateable moment recorded in a source. Births, marriages, deaths, enumerations, probate, court appearances.	Enumeration date: 24 April 1921.
Fact	A specific claim made by a record about a person, place, or event.	Occupation recorded as Plumber & Electrician.
Field	A single column in a spreadsheet or database. Each field holds one type of information.	Forename_Recorded_As is a field. It holds exactly one piece of information per row.
Forename_Normalised	The standardised, modern spelling of a forename, separated from how it was recorded.	JAS normalised to James.
Forename_Recorded_As	The forename exactly as it appears in the source, including abbreviations, misspellings, and variant forms.	Eliz / Thos / Jno / Wm / JAS
Identity_Evidence	The specific source or sources that support a decision about who a person is.	Cross-referenced with baptism register Portsmouth 1823 and will Bedfordshire 1827.
Identity_Status	A flag showing whether you have made a decision about who a person is. Three values: Confirmed, Unresolved, Provisional.	Unresolved means the question is still open. That is good practice, not failure.
Linked_Person_ID	The unique identifier of another record you have confirmed refers to the same individual.	P001 in the 1921 census links to P001 in the 1889 marriage register.
Linked_Source_Ref	The reference of another source you have connected to the same person.	CEN/1921/PORTSMOUTH links to MAR/STLUKES/1889/384
Normalised	Converted to a standard, consistent form. Separate from the recorded form. Never overwrites it.	01 Aug 1905 is the normalised form of 1/Aug/05.

Normalised_Reason	A note explaining why you normalised a name, date, or place in the way you did.	Thos normalised to Thomas — cross-referenced with signed will 1827 where full name appears.
Notes	Free text field for anything unclear, ambiguous, contradictory, or requiring follow-up research.	CRITICAL: Surname recorded as Fooke (one letter dropped) — retained in Recorded_As.
Person_ID	A unique identifier you assign to each individual in your dataset. Appears in every record linked to that person.	P001 is Eli Fookes. Every record for Eli Fookes carries P001.
Person_Role	The role a person plays in a particular record.	Head / Wife / Daughter / Boarder / Testator / Executor / Witness / Beneficiary / Groom / Bride
Place_Normalised	The standardised modern form of a place name.	St Mary, Portsea, Portsmouth, Hampshire.
Place_Recorded_As	The place name exactly as written in the source.	Wingfield in the Parish of Malgrave
Provisional	A working assumption about identity that has not yet been confirmed by independent evidence.	Provisional — believed to be same individual as P003 but awaiting corroboration.
Recorded_As	The principle of recording source material exactly as it appears before any interpretation or normalisation.	Albert James recorded as Albert James even though all other records give Adam.
Repository	The institution or archive that holds the original source.	The National Archives / Portsmouth History Centre / Bedfordshire Record Office
Row	One line of data in a spreadsheet. In wide format, one row = one record. In multi-line format, one row = one person in one record.	The census household of six people produces six rows in multi-line format.
Source	The original document or record you are working from. Not a summary, not a transcript, not someone else's interpretation.	The 1921 census schedule signed by Eli Fookes is the source.
Source_Ref	A unique reference you create to identify each source consistently across your dataset.	CEN/1921/PORTSMOUTH/GLADYS or WILL/BEDS/1827/024
Source_Type	The category of document.	Census / Parish Register / Will / Court Record / Newspaper / Electoral Roll / Military Record
Structured data	Information organised into consistent, labelled, separable parts that can be queried, compared, and linked.	A spreadsheet with named columns, one piece of information per field, is structured data.
Surname_Recorded_As	The surname exactly as it appears in the source, including variant spellings.	Fooke (not Fookes) as recorded in the 1889 marriage register.
Unresolved	An identity or claim that has not yet been decided. The question is still open.	Identity_Status: Unresolved — two records with matching name and approximate age but insufficient evidence to confirm.