

How to Write Structured Data Field Labels

How to write field labels, how to use the *Recorded_As* and *Normalised* pair correctly, how to record the reason for a normalisation decision, and how to build a consistent identity and source reference system.

Written for the complete beginner. No technical knowledge assumed. If you are not sure how to construct a label or what to do when two records contradict each other, this is the sheet to read first.

THE BASIC RULES	
Use underscores, not spaces	Field names cannot contain spaces. Use an underscore every time you would use a space. Correct: Surname_Recorded_As Wrong: Surname Recorded As Wrong: SurnameRecordedAs
Capitalise each word	Start each word with a capital letter. This makes labels readable at a glance. Correct: Place_Recorded_As Correct: Identity_Status Wrong: place_recorded_as Wrong: PLACE_RECORDDED_AS
One concept per field	Each field should hold exactly one piece of information. If you find yourself writing two things in one field, split it. Correct: Forename_Recorded_As and Forename_Normalised as separate fields Wrong: Name (holding both forename and surname together)
No commas in field names	Commas are used in CSV files to separate fields. A comma inside a field name breaks the file. Correct: Fact_1_Value Wrong: Fact, Value
Keep labels short but clear	Four words maximum. If you need more words to explain the field, use the tooltip row or the glossary. Correct: Place_Birth_Recorded Correct: Identity_Evidence Too long: The_Place_Where_The_Person_Was_Born_As_Written
THE RECORDED_AS / NORMALISED PAIR	
Always use both fields together	Every piece of recorded information should have two fields: what the source says, and what you have standardised it to. Never overwrite one with the other. Source says: Thos Recorded_As: Thos Normalised: Thomas Normalised_Reason: Full name appears in signed will 1827
Always add a Normalised_Reason	Record why you made the normalisation. This is the most commonly missed field. Without it, your normalised data is just an assumption with no evidence. Wrong: Forename_Normalised = Thomas (no explanation) Correct: Normalised_Reason = Thos expanded to Thomas — cross-ref signed will ABPW1827-24 where full name appears

The Albert James problem	When a source gives a name that conflicts with every other record, do not normalise it away. Record it exactly. Flag it in Notes. Keep the Normalised_Reason empty until you have an explanation. Forename_Recorded_As: Albert James Forename_Normalised: Adam Normalised_Reason: All other records (birth, military, death) give Adam. This record anomalous — reason unknown. Identity_Status: Provisional
IDENTITY STATUS — THE THREE VALUES	
Confirmed	You have verified this person's identity against at least one independent source. Record what that source is in Identity_Evidence. Identity_Status: Confirmed Identity_Evidence: Cross-ref baptism St Mary Portsea 1823 — name, father, approximate age all consistent
Unresolved	You have not yet made a decision. The question is still open. This is not a failure. It is honest. Identity_Status: Unresolved Notes: Two John Smiths with matching approximate age in same parish. Insufficient evidence to distinguish.
Provisional	You have a working assumption but have not yet found corroborating evidence. Use this when you are fairly confident but cannot yet confirm. Identity_Status: Provisional Identity_Evidence: Name, age, and address consistent with P003 in 1881 census — awaiting further corroboration
PERSON_ID — YOUR LINKING SYSTEM	
Assign an ID to every individual	Give every person a unique identifier the moment you first record them. Use a simple prefix and number. Examples: P001 / P002 / P003 Or with surname prefix: FOOKS001 / FOOKS002 Consistency matters more than the format you choose. Pick one and use it throughout.
Use the ID to link records	When you find a second record for the same person, use their existing Person_ID. Add the new source reference to Linked_Source_Ref. Person_ID: P001 Linked_Source_Ref: MAR/STLUKES/1889/384, WILL/BEDS/1827/024 This is how you build connections without losing the independence of each source record.
When you are not sure	If you are not certain two records refer to the same person, create a new provisional ID. Do not assume. Do not merge. P001 — Eli Fookes (confirmed across three records) P001a — Eli Fookes (provisional — possibly same, awaiting evidence) Better to have two IDs that turn out to be one person than one ID that turns out to be two people.

SOURCE REFERENCES — HOW TO BUILD THEM

Build a consistent reference format

A Source_Ref is a unique label you create for each source. Use the same format every time. A simple approach:

TYPE/PLACE/YEAR/NUMBER

Examples:

CEN/1921/PORTSMOUTH/GLADYS

WILL/BEDS/1827/024

MAR/STLUKES/1889/384

BAPT/PORTSEA/1823/042

No spaces, no commas

Apply the same rules as field labels. Underscores or forward slashes are both acceptable. Pick one and use it throughout.

Correct: CEN/1921/PORTSMOUTH/GLADYS

Correct: CEN_1921_PORTSMOUTH_GLADYS

Wrong: Census 1921, Portsmouth, Gladys Avenue