



*Teaching
Genealogists AI™*
Instructor: Sylvia
Vargas

AI Tools for Research and Organization

Module 3 – Organizing Your Data with AI

Course Instructor: Sylvia Vargas
Organization: Puerto Rican Genealogy Group
Website: <https://puertoricangenealogy.org>
Date Created: 2026-03-05
Estimated Time: 25–35 minutes

Learning Objectives

By the end of this module, you will:

- Know how to use AI to extract structured data from transcriptions you provide
- Be able to standardize names, dates, and place names across a dataset

- Understand how to convert narrative text into GEDCOM-compatible data
 - Know how to connect AI-organized output to your research log
-

The Organization Problem

Genealogists accumulate data in many formats: PDF transcripts, handwritten notes, photos of documents, oral history recordings, snippets from books, and long-form notes from interviews with relatives. When you are researching a family across multiple record systems – Spanish colonial, U.S. federal, church, and civil – the same person’s name may appear in dozens of variations across dozens of files.

AI cannot retrieve records for you, but it is exceptionally good at taking unstructured text you provide and returning it in a consistent, organized format. This module covers three workflows: extracting data into tables, standardizing names and dates, and converting narrative to GEDCOM format.

Workflow 1: Extracting Data into Tables

The most immediate use of AI for data organization is turning a batch of transcriptions into a structured table. Instead of manually copying each field from each record, you paste the transcriptions and tell the AI exactly what columns you want.

Prompt template:

I have a set of [record type] transcriptions from [location], approximately [date range]. Please extract the data into a Markdown table with these exact column headers: [list your headers]. If a field is not present in a record, enter “not recorded” in that cell. Do not infer or add information that is not in the transcription. Flag any uncertain readings with [?].

[Paste transcriptions here]

Example – baptismal records from a parish book:

I have a set of baptismal record transcriptions from a parish book in Utuado, Puerto Rico, approximately 1870–1880. Please extract the data into a Markdown table with these exact column headers: Date of Baptism, Child’s Name, Father’s Name, Mother’s Name, Paternal Grandparents, Maternal Grandparents, Officiating Priest, Notes. If a field is not present in a record, enter “not recorded” in that cell. Do not infer or add information that is not in the transcription. Flag any uncertain readings with [?].

En 12 de junio de 1875 bauticé a Ricardo, hijo legítimo de Manuel Ortiz y Rosa Santiago. Abuelos pat: Pedro Ortiz y Julia Ramos. Mat: Jose Santiago y Maria Colon. Cura: P. Medina.

En 3 de agosto de 1876 bauticé a Luisa, hija legítima de Rafael Torres y Carmen Rivera. Abuelos pat: no constan. Mat: Francisco Rivera y Ana Vega. Cura: P. Medina.

What the AI returns:

Date of Baptism	Child's Name	Father's Name	Mother's Name	Paternal Grandparents	Maternal Grandparents	Officiating Priest	Notes
12 June 1875	Ricardo Ortiz Santiago	Manuel Ortiz	Rosa Santiago	Pedro Ortiz, Julia Ramos	Jose Santiago, Maria Colon	P. Medina	
3 August 1876	Luisa Torres Rivera	Rafael Torres	Carmen Rivera	not recorded	Francisco Rivera, Ana Vega	P. Medina	

Important: Review every cell against the original transcription. The AI may silently add a child's surname (as shown above – “Ricardo Ortiz Santiago” is inferred from the parents' names, not stated in the record). Flag any field the AI completed beyond what the transcription states.

Workflow 2: Standardizing Names, Dates, and Places

When you have a dataset with inconsistent spellings, date formats, and place names, AI can scan the full set and propose standardized versions for your review.

What to standardize:

- **Dates:** Convert “15 de agosto, 1900,” “Aug 15 1900,” and “8/15/1900” to a single format such as ISO 8601 (1900-08-15) or a format compatible with your genealogy software
- **Place names:** Identify that “Bayamon,” “Bayamón,” and “Pueblo de Bayamón” all refer to the same municipality
- **Name spelling:** Flag that “Jimenez,” “Jiménez,” and “Gimenez” may be the same surname and propose a standardized form
- **Given names:** Note that “María de los Ángeles,” “María de los Angeles,” and “M. de los Ángeles” appear to be the same person

Prompt template:

I have a dataset of genealogical records from Puerto Rico. Please review these entries and: (1) standardize all dates to YYYY-MM-DD format, (2) identify place name variants that likely refer to the same location and propose a standard form, (3) flag surname spellings that may be phonetic variants of the same name, (4) flag given name variants that may refer to the same person. Do not change any data – return a table showing the original value, your proposed standard form, and your confidence level (high / medium / uncertain).

[Paste dataset here]

Using the output: The AI's proposed standardizations are a starting point, not a decision. Review each suggestion, verify it against the original records or a reference source, and make changes to your own dataset deliberately.

Workflow 3: Converting Narrative to GEDCOM-Compatible Data

One of the most powerful uses of AI for organization is converting written family stories or compiled genealogy notes into structured GEDCOM data. GEDCOM (Genealogical Data Communication) is the standard file format used by FamilySearch, Ancestry, and most genealogy software.

AI cannot create a verified GEDCOM file – that requires you to confirm every fact against original sources. But it can convert a narrative into the structured data fields a GEDCOM file requires, which you then verify and import into your software.

Prompt template:

I am going to provide a written family history narrative. Please extract the genealogical data and return it as a structured list of individuals, with the following fields for each person: Full name, Birth date, Birth place, Death date, Death place, Father's name, Mother's name, Spouse's name, Marriage date, Marriage place, Children's names. Mark any field where the information is inferred rather than directly stated. Mark any field where the information is uncertain.

[Paste narrative here]

Example input:

Juan Rivera was born around 1855 in Rincón, Puerto Rico. He married Petra Morales sometime in the 1870s. They had at least three children: Luis, born ca. 1878; Carmen, born ca. 1881; and Rosa, born ca. 1885. Juan died before 1910, as Petra appears as a widow in the census that year. Petra's parents were Francisco Morales and Dolores Vega, both from Añasco.

Example output structure:

Individual: Juan Rivera

Birth: ca. 1855, Rincón, Puerto Rico [stated]

Death: before 1910 [inferred from widow status of spouse in 1910 census]

Father: not stated

Mother: not stated

Spouse: Petra Morales

Marriage: ca. 1870s, place not stated [inferred from birth years of children]

Children: Luis Rivera Morales (ca. 1878), Carmen Rivera Morales (ca. 1881), Rosa Rivera Morales (ca. 1885)

Individual: Petra Morales

Birth: not stated

Death: after 1910 [inferred]
Father: Francisco Morales [stated]
Mother: Dolores Vega [stated]
Birth place of parents: Añasco, Puerto Rico [stated]
Spouse: Juan Rivera
Children: Luis, Carmen, Rosa

Notice that the AI correctly flags inferences (death date inferred from widow status) and notes where information is not stated. Review every field before entering it into your genealogy software.

GPS Checkpoint

Where does GPS apply to organized data?

Organization is not a GPS element – it is a practical precondition for applying the GPS. But how you organize your data has GPS consequences:

- **Source citations must survive organization.** When you extract data into a table, the source citation for each row must be preserved. A fact without a citation is unverifiable.
- **Inferences must be labeled.** When AI infers a child’s surname from parents’ names, or infers a death date from a widow’s status, that inference must be labeled as such. Unlabeled inferences become unverifiable facts.
- **Standardization is not transcription.** When you standardize “Bayamón” from “Bayamon,” you are making an editorial decision. Note it.

Practical rule: Add a “Source” column to every extraction table. Never let a row exist without a citation linking it to the original record image.

Connecting AI Output to Your Research Log

AI-organized data is most useful when it connects directly to your research log. After any AI extraction session:

1. **Save the raw AI output** – paste it into your notes before editing anything
2. **Add source citations** – link every row back to the original record
3. **Flag AI inferences** – mark anything the AI added beyond what the source states
4. **Note what you verified** – record which cells you confirmed against the original image and which are still pending

If you use a spreadsheet as your research log, add columns for “Source,” “AI-inferred (Y/N),” and “Verified against original (Y/N).” These columns protect you from treating organized data as confirmed facts.

What's Next

In Module 4 – Research Planning: Breaking Through Brick Walls, you will learn how to use AI to audit your research timeline, identify gaps, generate hypotheses, and build a structured research plan for ancestors who disappear from the record.

[<- Module 2](#) · [Back to Course Overview](#) · [Module 4 ->](#)

© 2026 Sylvia Vargas. Teaching Genealogists AI™. All rights reserved.