Metadata Schema for Project Vox Philosopher Networks Database

Karen Nielsen, Junyi Tao, and Emilie Menzel

Note: While this schema outlines how to input correspondence, this database method is not limited to correspondence. It is designed to handle a variety of connections such as family connection, birth date, death date, visit, etc. The full picture is about people's connection whatever those may be. It is our hope that in the future this schema can be expanded to document and visualize a variety of interpersonal connections.

A. Project Vox Philosopher Correspondence

For connection type as correspondence, follow these guidelines.

I. About the Schema

1. Audience and Purpose

This schema is to be used by the Project Vox research team for documenting letter correspondences by, to, and about a specific Project Vox philosopher. The schema will allow for a standardized data collection process for correspondences. By following this schema, researchers will also be able to input their research directly into a database that can allow for easily generated data visualizations.

2. Updates from Original Schema

This schema has been created as a standardized version of the research team's correspondence spreadsheets, particularly drawing from that documenting Princess Elisabeth's letter correspondence with contemporary philosophers. Karen Nielsen and Junyi Tao, for 2022 Story+, standardized the date format and translated the names into a controlled vocabulary. They also separated the multiple attribute category "To/From & Location" into distinct attributes and added attributes for thematic, people, and text subject headings. They removed the attribute "Philosophical Significance," as guidelines for deciding significance were unclear and the column was often left blank.

3. Domain

A letter counts as an individual item. A single letter is bookended by a salutation and signature and may include sections within which were written across a period of time. For example, on October 10, 1646, Princess Elisabeth wrote to Descartes "I have so little leisure to write here that I am constrained to send you this draft, in which you can see from all the difference in pens all the times I have been interrupted." Because the parts are bookended by a single salutation and a single signature, these parts are to be considered as only one letter.

A letter's reply is another individual item. For example, on May 6, 1643 Princess Elisabeth wrote a letter to René Descartes. This is one item. Then on May 21, 1643 René Descartes wrote a reply. This reply letter is another individual item.

Also to note, letters that are written in the same time frame but have their own salutations and signatures are counted as individual items. For example, we have two letters that René Descartes wrote to Princess Elisabeth in close succession during May of 1646 (no date of the month is available). Each letter includes its own salutation and signature. Each letter is thus considered an individual item.

4. Scope

This template was developed for letters containing philosophical themes written during the early modern period that involve philosophers on the Project Vox website. These letters are those commonly recognized as letter correspondence by the philosophy academic community. The scope is correspondence by, to, and about a specific Project Vox philosopher.

5. Differentiation

Letters are distinguished by their correspondence_id. At the time of record into the spreadsheet, every letter is to be given a unique id using the following guidelines:

[Project Vox Philosopher Initials] + . + [ascending chronological number]

- Each id begins with the initials of the relevant Project Vox philosopher. For example, all Princess Elisabeth correspondence correspondence_ids begin with "PE."
- This is to include a number

II. Attributes

1. connection_type
   a. Definition: The primary name for the medium in which the project vox philosopher and the correspondent are interacting.
   b. Guidelines & Best Practices: This attribute is included for hopes that in the future Project Vox will be able to collect data about a multitude of interactions. However, at this point this metadata schema is limited to recording correspondences.
   c. Repeatable / Not Repeatable: Not Repeatable
   d. Mandatory / Not Mandatory: Mandatory
   e. Data Type: String
   f. Formatting: Controlled Vocabulary
      i. Correspondence
   g. Examples: Correspondence

2. correspondent
    a. Definition: The Project Vox philosopher featured in or relevant to the connection.
    b. Guidelines & Best Practices: Use one consistent name for the correspondent. Use Library of Congress Name Authority File names whenever possible. If LCNAF is inappropriate or unavailable, you may use an LCNAF Variant, the name as used by the Stanford Encyclopedia of Philosophy, or create a new naming using LCNAF naming conventions. If using anything other than the LCNAF primary name, record the name and convention used in the Project Vox Airtable database..
    c. Repeatable / Not Repeatable: Not Repeatable
    d. Mandatory / Not Mandatory: Mandatory
    e. Data Type: String
    f. Formatting: Names. Use LCNAF formatting conventions.
    g. Examples:
        i. Descartes, René, 1596-1650 *(source: LCNAF)*
        ii. Elisabeth, Princess of Bohemia, 1618-1680 *(source: Stanford Encyclopedia using LCNAF date conventions)*

3. date_as_written
    a. Definition: point in time as recorded on the correspondence
    b. Guidelines & Best Practices: Write the date how the date is represented in your source. If you have more than one source, specify which one's date you are using in the note section.
    c. Repeatable / Not Repeatable: Not Repeatable
    d. Mandatory / Not Mandatory: Not Mandatory
    e. Data Type: String
    f. Formatting: Follow Society of American Archivists guidelines:
        i. record exact dates in a consistent and unambiguous fashion, expressed as year-month-day
        ii. https://saa-ts-dacs.github.io/dacs/06_part_I/03_chapter_02/04_date.html
    g. Examples:
        i. 1643 May 6
        ii. 1668 June/July, 1668 (meaning: June *or* July of 1668)
        iii. 1668 June-July (meaning: June to July of 1668)

4. reformatted_date
    a. Definition: A standardized date formatting that the computer can read.
    b. Guidelines & Best Practices: If all parts of the date are known, enter then as MM/DD/YYYY. If a day of the date is unknown, put questions marks as placeholders. If the date of the month or year is known to be within a certain frame, create all possible frames by creating repeatable dates.

      c. Repeatable / Not Repeatable: Repeatable
          i. If Repeatable, separate values by a comma.
      d. Mandatory / Not Mandatory
      e. Data Type: Date
      f. Formatting: Date
      g. Examples:
          i. Date: January 9, 1670
             1. reformatted_date: 01/09/1670
          ii. Date: Summer 1653
             1. reformatted_date: 05/??/1653, 07/??/1653, 08/??/1653

5. author
      a. Definition: the writer of the letter
      b. Guidelines & Best Practices: Use one consistent name format for the correspondent. Use Library of Congress Name Authority File names whenever possible. If LCNAF is inappropriate or unavailable, use an LCNAF Variant or the name as used by the Stanford Encyclopedia of Philosophy
      c. Not Repeatable
      d. Mandatory
      e. Data Type: String
      f. Formatting: Use LCNAF formatting conventions.
      g. Examples:
          i. Descartes, René, 1596-1650
          ii. Elisabeth, Princess of Bohemia, 1618-1680

6. recipient
      a. Definition: the receiver of the letter
      b. Guidelines & Best Practices: Use one consistent name format for the correspondent. Use Library of Congress Name Authority File names whenever possible. If LCNAF is inappropriate or unavailable, use an LCNAF Variant or the name as used by the Stanford Encyclopedia of PhilosophyIf the person goes by multiple names, choose one to be the correspondent name, and list other names under the notes section. Use one consistent name format for the name whenever it appears (from, to and correspondent should be one name if it is referring to one person).
      c. Repeatable / Not Repeatable
      d. Mandatory / Not Mandatory
      e. Data Type: String
      f. Formatting: Use LCNAF formatting conventions.Names
          i. When there is a first name and a last name, list name as last name, first name.

      g. When there is a title associated with a name, write the name as it is presented in the source.

      h. Examples:

          i. Descartes, René, 1596-1650

          ii. Elisabeth, Princess of Bohemia, 1618-1680

7. location

      a. Definition: A location given in the salutation of the letter.

      b. Guidelines & Best Practices: Typically the location refers to where the writer wrote the letter. If it is referring to a different location, add that in the notes section. If the letter includes no location, leave this attribute blank.

      c. Repeatable / Not Repeatable: Not Repeatable

      d. Mandatory / Not Mandatory: Not Mandatory

      e. Data Type: String

      f. Formatting: Free Text

      g. Examples:

          i. The Hague

          ii. Berlin

8. full text

      a. Definition: Record whether the full text of the correspondence is available. Make sure that the full text is found at the source.

      b. Guidelines & Best Practices: If the full text is not found at the source you are using, record where the full text can be found in notes. If the full text is not available, write No.

      c. Repeatable / Not Repeatable: Not Repeatable

      d. Mandatory / Not Mandatory: Mandatory

      e. Data Type: Boolean

      f. Formatting: Yes or No

      g. Examples:

          i. Yes

          ii. No

9. source

      a. Definition: The resource from which you, the researcher, are recording this information.

      b. Guidelines & Best Practices: MLA citation

      c. Repeatable / Not Repeatable: Repeatable

          i. If repeatable, separate with a semicolon.

      d. Mandatory / Not Mandatory: Mandatory

      e. Data Type: String

    f. Formatting: MLA citation

    g. Examples:

10. themes
   a. Definition: thematic content tags of the letter
   b. Guidelines & Best Practices:
      i. Choose recognizable topics that could lead someone to use this letter as a source. Imagine it as if you were making an index for the content of this letter, what tags would you give it?
   c. Repeatable / Not Repeatable: Repeatable
      i. If Repeatable, separate with a semicolon.
   d. Mandatory / Not Mandatory: Not Mandatory
   e. Data Type: String
   f. Formatting: Free Text
   g. Examples: Empiricism; Rationalism; Stoicism

11. people
   a. Definition: The name of a person (not the sender or receiver) who is referred to or explicitly mentioned in the text.
   b. Guidelines & Best Practices: If a name is mentioned in the footnotes as being referred to, include it in the people section of this database.
   c. Repeatable / Not Repeatable: Repeatable
      i. If Repeatable, separate with a semicolon.
   d. Mandatory / Not Mandatory: Not Mandatory
   e. Data Type: String
   f. Formatting: Free Text
   g. Formatting: Names
      i. Last Name, First Name
      ii. Official title as presented in the text
   h. Examples:
      i. Queen of Sweden

12. Works
   a. Definition: another piece of writing (a "work") that's being referred to or explicitly referenced within the letter
   b. Guidelines & Best Practices: If a text is mentioned in the footnotes as being referred to, include it in the text section of this database. If the text is being referred to as a specific translation or edition, include that as part of its name.
   c. Repeatable / Not Repeatable: Repeatable
      i. If Repeatable, separate with a semicolon.
   d. Mandatory / Not Mandatory: Not Mandatory
   e. Data Type: String
   f. Formatting: Free Text

g. Formatting: Free Text
    i. Write the text as it is referred to in the original letter
    ii. Examples:
        1. Principles of Philosophy

13. notes
    a. Definition: Anything additional to the above attributes you think is important to include or 'note' about the letter, and anything that has been designated in the above attributes to be included in the notes section (i.e. alternate names).
    b. Guidelines & Best Practices: It is always best to record when you make a choice or see something that is unclear.
    c. Repeatable / Not Repeatable: Repeatable
        i. If Repeatable, separate with a semicolon.
    d. Mandatory / Not Mandatory: Not Mandatory
    e. Data Type: String
    f. Formatting: Free Text
    g. Examples:
        i. Princess Elisabeth is referred to as Princess Elizabeth in this letter
        ii. Half of this letter is missing

14. agent
    a. Definition: The name of the person recording data in the spreadsheet. This will not be viewed by the public; it will only be kept internal to Project Vox records.
    b. Guidelines & Best Practices: Make sure you add in your name each time you add in data. If you are editing someone else's entry, add in yourself as the next editor.
    c. Repeatable / Not Repeatable: Repeatable
    d. Mandatory / Not Mandatory: Not Mandatory
    e. Data Type: Name
    f. Formatting: Names
        i. Last name, First name
    g. Examples:
        i. Nielsen, Karen
        ii. Tao, Junyi

15. input_date
    a. Definition: The date at which the inputter adds in data to the spreadsheet. This will not be viewed by the public; it will only be kept internal to Project Vox records.
    b. Guidelines & Best Practices: Try to add in the date when you add in your name. You can copy-paste your name and date to all the entries you worked on in one setting.
    c. Repeatable / Not Repeatable: Repeatable
    d. Mandatory / Not Mandatory: Not Mandatory

e. Data Type: Date
  f. Formatting: Date
      i. MM/DD/YYYY
  g. Examples:
      i. 11/06/2022

-------------------------Guideline's from Emilie about recording attributes----------------------------

For every attribute (i.e. column in the spreadsheet), include the following:

1. *Name of Attribute:* **what are you calling this attribute?**
    a. Be sure to copy any specific formatting, capitalization, and spacing you are using in the database. First Name, First_Name, firstName, and FirstName are each distinct attributes.
    b. e.g. WitchName
    c. e.g. AlternateNames

2. *Definition of Attribute:* **what content is recorded here?**
    a. Briefly describe what the attribute is. An attribute is the category (e.g. WitchName), and a value is what is entered for the attribute for an individual item (e.g. Baba Yaga).
    b. e.g. "The primary name by which stories including this character refer to the witch."
    c. e.g. "Additional names by which stories including this character refer to the witch."

3. *Guidelines & Best Practices:* **how do you determine the value for the attribute?**
    a. Here's where you address common confusions you think the researcher might encounter when choosing a value (e.g. Baba Yaga) to fill in for the attribute (WitchName). Offer instructions for how the researcher should make decisions about what to record.
    b. e.g. "If character goes by multiple names, choose one to be the WitchName and list others under AlternateNames. You may choose one as a primary name for several reasons, including: it is widely recognized, it is most commonly used, it is the name of the witch in the native culture's language. Witch's name may be general, a phrase or an epithet. A story may only refer to witch as "the witch," "the old woman," or "the hag," and these are acceptable entries for Witch Name. Witch's name (and alternate names) may include what the witch calls themself or what others call the witch."

4. *Repeatable / Not Repeatable:* **can there be multiple values for this attribute?**
    a. Is it ok to have more than one answer for this attribute? If so, the attribute is Repeatable. If there needs to one value selected and only one, the attribute is Not Repeatable.
        i. *If Not Repeatable:* In the guidelines section, explain to the researcher how to choose between possible options.

       ii.   *If Repeatable:* separate values by a semicolon. Give instructions if there's a particular order in which the values should be entered.

  b. e.g. WitchName: Not Repeatable. "If character goes by multiple names, choose one to be the Witch Name and list others under Alternate Names. You may choose one as a primary name for several reasons, including: it is widely recognized, it is most commonly used, it is the name of the witch in the native culture's language."

5. *Mandatory / Not Mandatory:* **does the researcher have to include a value here?**
   a. Is it ok for the researcher to not record anything for this attribute (i.e. leave the value null)? Can the box be left blank? If so, the attribute is Not Mandatory. If the researcher should always put some value down for this attribute, the attribute is Mandatory.
   b. Attributes used for uniquely identifying items are usually mandatory.
   c. e.g. WitchName: Mandatory
   d. e.g. AlternateNames: Not Mandatory

6. *Data Type:* **how should the computer interpret the value of this attribute?**
   a. Categorize the format of information being shared by this attribute.
   b. **Common Options:**
      i. **Integer:** values entered for this attribute should only be whole numbers
      ii. **String:** values entered for this attribute can be a sentence, a word, a phrase, text, or another string of symbols.
      iii. **Date:** Date in the YYYY-MM-DD format. "December 25, 2022" would be categorized as a string. "2022-12-25" would be categorized as a date.
      iv. **Boolean:** use when there are only two answer options for this attribute, usually yes or no.
   c. e.g. WitchName is a String

7. *Formatting:* *how should the researcher write up the information in this category?*
   a. Give directions for how the researcher should format the content for this attribute. How should the content be presented?
   b. **Common Options:**
      i. **Free Text:** anything goes!
      ii. **Controlled Vocabulary:** the researcher should use the formatting and phrasing as designated by a predetermined vocabulary list. The researcher needs to choose from the options available on the list, copying the formatting and phrasing exactly.
         1. Supports searchability and consistency
         2. Frequently used for subject headings and names
         3. Examples of controlled vocabularies include Library of Congress Subject Headings, Library of Congress Name Authority File, and The Getty Art and Architecture Thesaurus
      iii. Names: e.g. format as Last Name, First
      iv. Dates: e.g. should it be YYYY-MM-DD? or YYYY-DD-MM?
   c. e.g. WitchName: Last, First: Yaga, Baba or Smith, John

8. ***Examples:*** It's often helpful to provide some examples for what your guidelines ought to look like in practice