

Conventions for Establishing xml:id Values

People

The editors expressed a desire for identifiers to be human readable rather than alphanumeric. The conventions below build off those used for @xml:id values in the [Map of Early Modern London Project](#)

1. Take the first four letters of the individual's last name and retain the existing case.
 - a. If there are less than four letters, use all the letters (e.g. just "Fox" for "Fox, C. J.")
 - b. Typically only the first of these four letters will be capitalized, but for derived names or bicapital nouns more than one letter will be capitalized. If a letter is capitalized in the last name, keep it that way for the ID (e.g. "OCon" for "O'Connor, Don").
2. Add a capital first initial to the base ID you created in step 1
 - a. Ignore any honorifics or titles (e.g. "Lord", "Lady", "Count", "Mrs.", "Mr.")
 - b. If the first name is not known, insert the letter "X"
 - c. If only a last initial is known, insert the initial followed by "xxx".
3. Add a number to make the ID unique
 - a. Typically "1" needs to be added to the end. If the base ID is not unique, the number can be used to differentiate it (e.g. "EdgeR1" and "EdgeR2")
 - b. A check for uniqueness will be run on IDs by faceting the ID column in OpenRefine

Edge Cases

In some situations, no name at all may be known/decipherable, but we can trace this individual as a unique entity through the course of the letters. In this case use "Anon" with a digit (e.g. "Anon1") to uniquely identify this unnamed individual.

Examples

Name	ID	Rule
Scott, Sir Walter	ScotW1	Ignore titles/honorifics
McGrigor, James, Sir, 1771-1858	McGrJ1	Retain capitalization in last name
Moutray, Mrs.	MoutX1	Use "X" when first name is unknown
18th-c author, dictionary compiler	Anon1	When no actual name is known, use "Anon" for

		Anonymous and iterate the number value
D, Lady	DxxxX1	When the full last name is not known, add “xxx” after the last initial
Henry VIII., King of England 1491-1547	Henry8	When a first and last name are not present for royalty etc., use the first five letters of the name and translate the roman numerals to a number
Molé, Louis-Mathieu, comte, 1781-1855	MoleL1	When diacritics are present in a name, do not include them in the xml:id
Molière, 1622-1673	Molie1	For pseudonyms, use the first five letters of the pseudonym and add a number after it
Penelope	XxxxP1	When you have a first name only, follow the regular convention of using Xs.
Roy, General	RoyX1	When someone’s last name is less than four letters, include the last name without additional x values to account for the typical four characters

Places

As with people, established authorities were not able to be identified for all places mentioned in the letters. Initially, using the IDs already established in existing authorities for @xml:id values was considered. Here’s an example placeography with codes from the Michael Leary Letters - <https://www.tapasproject.org/tei-learning-docs-bc/files/leary-placeography>

1. Enter the specific place name with hyphens between individual words (if the place name is more than a single word)
2. Follow the specific place name with an underscore and the appropriate country code from the [MARC Country Code List](#)

Exceptions

1. If the place name is a country by itself, enter the 3-character code from the [MARC Country Code List](#) only
2. If the place name is located in a country with states (US or Australia) include a code for the state. For the US we will use postal code values as these are easily recognizable. We also don't need the trailing "U" for US state codes present in the MARC Country Codes for uniqueness. For other states, follow the MARC Country Code List.

Examples

Place Name	ID	Rule
England	ENK	If a country is the place name, enter the MARC Country Code as the ID
Carrickfergus	Carrickfergus_IE	Do not differentiate between Ireland and Northern Ireland in terms of codes (since Northern Ireland wasn't established in Edgeworth's time)
Boston	Boston_MA_US	Use a code for states when dealing with a country that has states. Note that an underscore is used following the specific place name.
Château de Champlâtreux	Chateau-de-Champlatreux_FR	Ignore diacritics. Also retain capitalization present in the controlled form of the name in Geonames. When a place name is composed of multiple words, separate the initial place name with hyphens. An underscore can be used between the initial place name and the country code.
St. Paul's Churchyard	St-Pauls-Churchyard_ENK	If the place name includes an abbreviation with a period, omit the period. If the place name includes an apostrophe, omit the apostrophe.
Corn Hill	Corn-Hill_XX	If unable to determine the country the place name is within, follow the name by

		“_XX”
--	--	-------

Works

As with people and places, established authorities were not able to be identified for all works mentioned in the letters. When possible, wikidata links are included to identify “works.” Works include any published content mentioned in the letters - novels, journal publications, essays, short stories, etc. Unlike with people and places, no examples of an existing “ography” file for works has been identified to use as a model.

1. Use the first five letters of the title (in English) followed by a number. Exclude initial articles (A, The, etc.) when establishing the @xml:id
2. If the title is only represented in a non-English language, use the first five letters of that title and a number to establish the @xml:id. Still be careful to exclude initial articles (Les, Des, etc.)
3. If the work is part of a larger work (short story or essay in a published volume), establish the xml:id using the same conventions. The only difference is that the @xml:id will be entered in a “Subwork” column on the Taxonomy spreadsheet on the same row with its associated parent work rather than directly in the main “ID” column.

Examples

Title	ID	Rule
The Absentee	Absen1	If a country is the place name, enter the MARC Country Code as the ID
Des Circonstances Actuelles qui Peuvent Terminer la Révolution et des Principes qui Doivent Fonder la République en France	Circo1	When the title is only given in a non-English language, use the first five letters of that title and exclude any initial articles
The Robin's Petition	Robin1	Subworks (short stories, essays, etc.) get individual xml:ids following regular conventions, but the @xml:id is listed in a separate “Subwork ID” column under the parent work rather than in a separate row with the xml:id entered in the “ID” column. See the Works tab of the Taxonomy spreadsheet for more information.

