What's in a Name?

Karen Coyle January 26, 2022

From https://kcoyle.blogspot.com/2022/01/whats-in-name.html

This is an essay about the forms of names and their representation in metadata. It is not by any means complete, nor am I an expert in this very complex area. These are my observations and a few suggestions for future work. All comments welcome.

If you do anything online, and surely you do, you have filled in countless forms with your name and address. Within the Western and English-speaking world, these have some minor (and occasionally annoying) variations. You might be asked for a first name and last name, or a given name and a family name, or just a name in a particular order.

Full name (First and Last name)
karen coyle

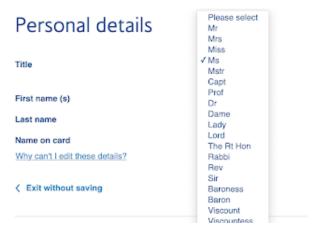First name
Karen

Last name
Coyle

There are variations, of course. Some recognize the practice of giving a person a "middle" name, that is a second, and perhaps secondary, additional name.

* First Name: | Middle Name: | * Last Name:

Because these forms are often used in commercial sites and the companies wish to have a polite relationship with their customers, you might be asked about your preferred form of address.

Dr.

Miss

Mr.

Mrs.

These forms of address have cultural significance, and the list itself can reveal quite a bit about a culture. This is the list from the British Airways site:

Personal details

Title

First name (s)

Last name

Name on card

Why can't I edit these details?

‹ Exit without saving

Please select
Mr
Mrs
Miss
✓ Ms
Mstr
Capt
Prof
Dr
Dame
Lady
Lord
The Rt Hon
Rabbi
Rev
Sir
Baroness
Baron
Viscount
Viscountess

We'll come back to some of these below.

The above examples come from commerce sites. The use of names at those sites are mostly social. Even on a site like a bank, the name has only a minor role in regards to identification because security relies on user names, passwords, and two-factor identification. Names themselves are poor identifiers because they are far from unique across a population. Even if you think you have an unusual name, you will find others with your name in the vastness of the Internet.*

If you think about times that you've been on the phone with some bank or service, they invariably ask you to provide a telephone number, an email address, or a unique identifier like a social security number as a way to identify you. Only after they have located a record with that identifier do they use your name as both a confirmation that you've given (and they've entered) the correct number, but also so that they can cheerily refer to you by your name.

Names in Cultural Heritage

Where commercial organizations use names to effect a relationship with their current customers, cultural heritage institutions have a different set of needs. They often cover not only names of modern persons but persons worldwide and of previous eras. An organization must be able to encode this full range of names in a way that is useful today but that is, to the extent possible, faithful to the cultural and historical context of the person. Royalty, religious figures, even characters in mythology all have a very tender place in their respective cultures. To treat them otherwise is to dismiss their cultural importance. You wouldn't want to provide metadata for Benedict XVI without also including that his title and his role in the church is "Pope". You most certainly would not simply name him "Joseph Ratzinger" unless you were giving a very specific, pre-Pope, context. I don't know what name Queen Elizabeth II would provide when signing up for an Amazon account ("Elizabeth Windsor"?) as there is unlikely to be an input box appropriate for her royal name, but culturally and historically she is Elizabeth II, Queen of Great Britain.

There is also the question of giving people their due rank in whatever hierarchy the particular culture values. As you can see with the example above for the list of the titles offered by British Airlines, whereas US-based airlines limit the titles to Mr., Mrs., Ms. and Dr., that titles of nobility are important in the UK. We can presume that to "mis-title" a person would be a social *faux pas* in most cultures, but there is also a historical context included in titles that one would not want to lose.

The "firstname, lastname" Problem

Not all names fit the "firstname, lastname" model. A primary reason to identify these parts of names is to support displays in alphabetical order by the "last name". This assumes that the last name is a family name, and that common usage is to gather together all persons with that family name in a display. In reality, this singular "family name" is only one possible name pattern.

As the term "family name" implies, this positions a person within a group of persons with a particular relationship. In the dominant Western world, the name is paternal and denotes a line of inheritance. But this is by no means the only name pattern that exists. There are cultures where the child's name includes the family names of both the mother and the father, and sometimes other ancestors in the family line. This is how Juan Rodríguez y García-Lozano and María de la Purificación Zapatero Valero, have a son named José Luis Rodríguez Zapatero. Treating "Rodríguez Zapatero"

as the family name would not bring together the alphabetical entries of the father and son.

There are other cultures that have a given name and a patronymic. While a patronymic may look like a family name, it is not. The singer Björk may have seemed to be using a single name as part of her art, like Cher or Madonna, but in fact in the Icelandic culture persons are known by a single name. When a more precise designation is needed, that name is enhanced with a name based on the given name of their father. In this case, Björk has a more "official" name of Björk Guðmundsdóttir, which is "Björk daughter of Guðmundur". Her father's name was Guðmundur Gunnarsson, he being the son of Gunnar. The author Arnaldur Indriðason is "Arnaldur the son of Indriði." In this practice, creating an order based on the patronymic would result in just a jumble of individual parental names, and persons are almost always called solely by their "first-and-only" name.

Yet another exception to the firstname/lastname conundrum relates to the names of royalty as mentioned above. Charles, Prince of Wales is the son of Elizabeth II. Their names do not connect them which is somewhat ironic given how important family relationships are to royal lineage. Both are of the house of Windsor but you wouldn't know that from their names. Like a Pope, the cultural or political position in these cases outweighs the personal. In addition, the title by which someone is officially known can change over time, making identification even more confusing, with titles being inherited or bestowed as circumstances change. Some people hold a plethora of titles: in addition to Prince of Wales, Charles is Earl of Chester, Duke of Cornwall, Duke of Rothesay, Earl of Merioneth and Baron Greenwich. This is as bad as the name proliferation in Russian novels, and just as confusing.

And there are the "one name" instances. We have historical figures with only a single name ("Homer", "Aesop") but there are also current cultures in which members have only one name.

> While praising the president's new constitution for giving Tamils a better deal, Thirumahal, a Tamil law student who has only one name, said she, too, felt it would not end the war.

> He said the authorities in New Delhi expected Mr. Laldenga, who has only one name, to "work within the Constitution." No special security precautions were being enforced in the remote state of roughly half a million people, he added.

Any metadata that strictly requires both a given name and a family name will be unable to accommodate these and it is not unusual for people with only one name to be required to provide a second name to conform to the given/family name expectation in other cultures. There may even be local traditions for how one invents such a name. Yet they would not use that invented name in their own home country.

Names and Language

It is hard to separate language from culture, but there are some name situations in which the name is translated into the "receiving" language. Catherine (the Great) is Catherine in French, Caterina in Italian, etc. The same is true of Popes:

> Papa Franciscus (Latin)

> Papa Francesco (Italian)

> Papa Francisco (Spanish)

> Pope Francis (English)

Another twist is that scientists and other cognoscenti of the late medieval and early modern times communicated with each other in Latin, and, probably as a form of showing that they were members of this elite club, often converted their names to a Latin form. Thus, one Aldo Pio Manuzio, a Venetian scholar and a very early book printer, took the name Aldus Pius Manutius. Francis Bacon published his "Novum Organum"

(which was in Latin) as "Franciscus Baconis".

Things get doubly complex as people and their names move from one culture to another. Many people of Chinese origin reverse the order of their names from family name first then given name to the preferred order in Western countries that places the family name last. In some cases, as with science fiction author Liu Cixin, a change for the Western marketplace creates a bit of confusion for anyone wanting to correctly encode this Chinese name.

Note that his translator, Ken Liu, an American, uses the Western form of his own name. So this book cover is a good illustration of the name problem across cultures.

Names in Metadata

How we handle names in metadata design depends mainly on the intended application functions for the data. I give below some key functions that use names, but this is an incomplete list. I can see these four as key purposes for names and their encoding in metadata:

1. **Display** - Names get displayed in a number of different contexts, from phone books to faculty listings on a web site to conference name tags. Displays may use all or only part of a name, and there are a variety of ways that one can order the name parts.

2. **Disambiguation** - Which Mary Jones is this? How do I identify and find the one that I am looking for?

3. **Addressing** - We do want to address people appropriately, and we also want to talk *about* them appropriately.

4. **Finding** - Searching via keyword is without context, so I'll assume that all name forms can be searched in that way. I will describe "finding" as meaning a search for a specific, known name.

I'll trace these through some metadata schemas to illustrate the metadata capabilities one might have.

Library of Congress (and other libraries)

Libraries have been dealing with names and name forms for, well, forever; as long as there have been libraries. The set of rules for determining what name to enter for someone in the library catalog is many, many tens of pages long, and there are separate rules for personal names, corporate names, and family names. Yet library name practices have their limitations, in particular that names are entered as strings that are to be used to create a specific alphabetical sort order that begins with the surname, followed by a comma, and then the forename(s).

> Dempsey, Martin, 1904-
> Dempsey, Martin E.
> Dempsey, Mary.
> Dempsey, Mary A.

Display by family name works well for Western names with family names, but not for Eastern names that place the

family name first.

> Mao, Zedong

Following Chinese name practice his name would naturally be given as "Mao Zedong" because the family name is always given first. If one attempts to use the comma to revert names to their natural order, say from "Smith, Jane" to "Jane Smith" then you would also end up with "Zedong Mao" which is not correct in that cultural context. A culturally sensitive "natural order" display is not supported by this metadata.

The primary display form is the Western one of lastname-comma-first names, but there are exceptions for entry by forename, which is given specific coding:

> Arnaldur Indriðason, 1961-
> Homer

As I've shown in the Mao Zedong example, the encoding of name parts in library data does not provide what you might need to create other display forms. In the case of Arnaldur Indriðason, outside of the library need to alphabetize its entries, you may want to know that Indriðason is a patronymic if you intend to use the name to address the person as he would be addressed in his culture. The example of "Mao, Zedong" is lacking the information that this is a name in a culture that regularly refers to people with their surname preceding their given name (and without a comma). You would want to know that this should be rendered as "Mao Zedong" when used in that context.

As you can see in the examples above, the Library of Congress name practice goes beyond just the name and adds elements that are meant to inform and clarify. It includes dates (birth, death); titles and other terms associated with a name (Pope, Jr., illustrator); enumeration (II); and fuller form of the name, which fills in portions of the name that use initials ("Boyle, Timothy D. (Timothy Dale)"). Interestingly, the "III" in Pope Pius III is an enumeration, while the "III" in "John R Kennedy, III" is an "other term associated with a name." I'm going to guess that this primarily relates to the positioning of the "III" in the display. This illustrates a tension between identifying parts of the name and providing the desired display of those parts.

There is a another problem with "title and other terms" because it is a catchall element that doesn't distinguish between some very different types of data. The documentation lists:

- titles designating rank, office, or nobility, e.g., Sir
- terms of address, e.g., Mrs.
- initials of an academic degree or denoting membership in an organization, e.g., F.L.A.
- a roman numeral used with a surname
- other words or phrases associated with the name, e.g., clockmaker, Saint.

As you can see, some of these would display before the name in a "natural order" display:

- *Sir* Paul McCartney
- *Mrs*. Harriet Ward

While others display afterward:

- John Kennedy, *Jr*.
- John Kennedy, *III*

And some can be either or both:

- *Dr*. Paul Johnson, *DDS*
- *Dr*. Sophie Jones, *Ph.D., F.I.P.A.*

There is always the need to disambiguate between people with the same name. Some of these "other terms" work well in identifying a person:

> Boyle, Tom *(Professor)*

> Boyle, Tom *(Spiritualist)*

However, the clarification between identical names used most often in library name data is the dates of birth and death. These used to be included only when necessary to distinguish between identical names but the information is now included whenever it is available to the cataloger. This makes the dates an integral part of the name, much as the roman numerals of the names of Popes.

> Pius I, Pope, d. ca. 154

> Pius II, Pope, 1405-1464

Although perhaps once useful for the purpose of distinguishing otherwise identical names, the sheer number of people who are included in library catalogs has greatly limited the utility of these dates for disambiguation.

> Kennedy, John, 1919-1945
> Kennedy, John, 1921-
> Kennedy, John, 1926-1994
> Kennedy, John, 1928-
> Kennedy, John, 1931-
> Kennedy, John, 1931-2004
> Kennedy, John, 1934-2012
> Kennedy, John, 1939-
> Kennedy, John, 1940-
> Kennedy, John, 1947-
> Kennedy, John, 1948-
> Kennedy, John, 1951-
> Kennedy, John, 1953-
> Kennedy, John, 1956-
> Kennedy, John, 1959-
> Kennedy, John, 1963-
> Kennedy, John, 1965-
> Kennedy, John, 1973-
> Kennedy, John, -1988.

There is provision for alternate versions of names in library practice although these reside in a separate file and are not always linked to the primary name in library databases.

> Boyle, Thomas John
>   see: Boyle, T. C.

The library name practices, although probably the most detailed of any metadata name schemes, are not very generalizable; they serve one designated application, which is the alphabetical order of the entries in the library catalog.

Dublin Core

Dublin Core is absolutely minimal when it comes to names, as "core" implies. It provides only one property, dct:creator, without further detail. It also does not distinguish between persons and organizations: both can be coded as "creator" with an implicit class of Agent. Any further intelligence must be provided elsewhere in a metadata scheme that makes use of Dublin Core.

Dublin Core does allow for the value of the dct:creator property to be either a literal or an IRI or Bnode, and the encoding of the value of the IRI could be a more precise name form. Using an IRI could also be a method for providing a unique identity for the creator.

FOAF

The "Friend of a Friend" vocabulary is about people, their names, and some modern social connectivity: email address, web site, etc. FOAF has three name properties:

- foaf:name - which can be used to an entire name, undifferentiated in terms of types of name
- foaf:familyName & foaf:givenName - intended to be used together (but with no mechanism to enforce that) this allows an obvious separation between the names. How they would display is left to the applications that make use of them.

The foaf:familyName and foaf:givenName cover a limited set of name forms. In the context of many online sites this may suffice, especially where there is no enforcement of "real" names. Given that FOAF was developed for use within and between online social sites, it avoids the need for historical forms of names.

All of these are defined as taking literal values, which we know does not provide an unambiguous identity for a person. There are properties defined in FOAF under the "Social Web" rubric, such as an email address, that should serve to disambiguate persons in a particular social context. These are not, however, part of the name itself.

schema.org

The vocabulary schema.org was developed to provide "structured data on the Internet". (This is exactly the original impetus behind Dublin Core. How that went south, and what schema.org attempts to do instead, is beyond this post.) The vocabulary listed under the person schema is extensive, although only a few elements are directly related to names:

> sdo:familyName, sdo:givenName, sdo:additionalName

sdo:givenName, is defined as the "first name" and sdo:familyName, is defined as the "last name". sdo:additionalName, is "An additional name for a Person, can be used for a middle name". This latter is highly flexible but at the same time non-specific. It also creates some confusion in terms of the order of names for anyone whose name does not fit the exact "first-middle-last" pattern. As shown above, it's not totally uncommon to have more than one name that can fit into any of those particular buckets. Presumably the properties are repeatable, but they are defined with the singular term "name". It also does not clarify a display order.

> sdo:givenName "T."
> sdo:familyName "Boyle"
> sdo:additionalName "C."

Schema.org does have properties for both pre-name and post-name honorifics. The examples given for these are: sdo:honorificPrefix (Dr., Mrs.); sdo:honorificSuffix (M.D., PhD). These examples don't make it clear if it might be possible to encode:

sdo:givenName "Charles"
sdo:honorificSuffix "Prince of Wales"

or

sdo:givenName "Pius"
sdo:honorificPrefix "Pope"
sdo:honorificSuffix "II"

In any case it appears that this would not distinguish between the informal honorifics like "Esq." and those that are essential parts of the name such as titles of nobility. There also does not seem to be an obvious way to encode non-honorific suffixes, such as "Jr." or "III".

Without some strong guidance, it would be hard to know which of these properties would be used for the parts of a name like María de la Purificación Zapatero Valero. We'll see a possible solution to this with Wikidata, below.

Wikipedia

Wikipedia has probably millions of articles for people and therefore has to deal with the question of names. Their search does not distinguish between names and other article topics, and all are searched in left-to-right natural order in a drop-down box. Names are article titles just as any topic can be an article title.

There is no special coding of the name or parts of the name - it is simply a string of characters. Where more than one person has the same name article creators must add something to disambiguate the name which is usually done by adding an area of activity and perhaps a location associated with that activity:

John F. Kennedy
John F. Kennedy International Airport
John Kennedy (Louisiana politician)
John Kennedy Toole
John Kennedy (Scottish footballer)
John Kennedy Sr. (footballer)
John Kennedy (Australian musician)
John Kennedy, 6th Earl of Cassilis
John Kennedy O'Connor
John Kennedy (radio broadcaster)

Wikipedia also has a special type of page where topics that have common terms, including names, can be further defined.

John F. Kennedy (disambiguation)

From Wikipedia, the free encyclopedia

John F. Kennedy (1917–1963) was the 35th president of the United States.

John F. Kennedy may also refer to:

People   [edit source]

- John F. Kennedy Jr. (1960–1999), son of John F. Kennedy
- John F. Kennedy (Georgia politician) (born 1965), member of the Georgia State Senate
- John Fitzgerald Kennedy, man convicted of the murder of Thomas and Jackie Hawks
- John Francis Kennedy (disambiguation), several people

These pages allow an explanation to distinguish between people who share a name. It goes beyond the parenthetical phrases that are used to create unique article names for persons with the same name, and is much more human-friendly than the birth and death dates that library cataloging relies on. Yet while Wikipedia excels in disambiguation, its encoding for names is limited to a single property, "name", in the infobox for a person, although it also allows for honorifics and for alternative forms of the name.

```
| honorific_prefix   =
| name               =
| honorific_suffix   =
| image              =
| image_size         =
| alt                =
| caption            =
| native_name        =
| native_name_lang   =
| birth_name         =
```

Because the various Wikipedias are divided by language, there are properties for translations and transliterations of names, and it allows for name changes over the course of a person's life.

Wikidata

Wikidata began by extracting data points from the Wikipedia entries, primarily from the infoboxes, but has grown beyond that to a database of facts that is edited directly. Perhaps because it is massively crowd-sourced, a [long list of name properties](#) have been developed. In addition to the usual *given name* and *family name* there are terms like *demonym* (a name representing a place), *second family name in Spanish name, Roman cognomen* (ancient surname)*, patronym or matronym* (names representing the person's father or mother)*, first family name in Portuguese name,* and many others.

Also because it is crowd-sourced there should be no expectation that this list is complete or balanced. It most likely represents a modicum of self-interest on the part of participants.

Conclusion (?)

Any solution in this area needs to recognize that one size does not fit all. For some applications a single "name=[string]" will be sufficient and it would be seriously counter-productive to force those folks to engage in detailed encoding. Another barrier to detailed encoding is that few people have knowledge to encode the universe of name forms at a detailed level. Requiring metadata creators to make distinctions outside of their understanding would only result in error-ridden metadata. Better a blind single string than mis-coded details. Yet there will be applications and their metadata communities that can or must make use of the subtleties of name details that are not of interest to others.

Because of both the great variety of name forms and the variability of applications that make use of names, I recommend a metadata vocabulary that follows the principle of minimum semantic commitment. This means a vocabulary that includes broad classes and properties that can be used as is where detailed coding is not needed or desired, but which can be extended to accommodate many different contexts.

The trick then is to define broad classes that aid in defining semantics but do little restriction. Classes for things like "Agent", with subclasses for "Person", "Groups of Persons", and perhaps "Non-persons". Properties could begin with "name" which could be subdivided into any definable part of a name that people find useful. Further specificity can be provided by application profiles that define such requirements as cardinality or value types for the various properties. Applications themselves could contain rules for the displays that are needed for their use cases.

The challenge now is to find a standards group that is interested to take this on.

----

* With perhaps a few exceptions. I once heard [Lorcan Dempsey](#) opine that person's names would be much more useful if parents would just give their children unique names, "... like Lorcan Dempsey."