



Personal names around the world

About this article

This article has been reviewed by the W3C Internationalization Working Group and has gone through public review to make it as accurate as possible.

If you have comments, please let us know.

Updated 2011-08-17 19:22

QUESTION

How do people's names differ around the world, and what are the implications of those differences on the design of forms, databases, ontologies, etc. for the Web?

People who create web forms, databases, or ontologies are often unaware how different people's names can be in other countries. They build their forms or databases in a way that assumes too much on the part of foreign users. This article will first introduce you to some of the different styles used for personal names, and then some of the possible implications for handling those on the Web.

This article doesn't provide all the answers – the best answer will vary according to the needs of the application, and in most cases, it may be difficult to find a 'perfect' solution. It attempts to mostly sensitize you to some of the key issues by way of an introduction. The examples and advice shown relate mostly to Web forms and databases. Many of the concepts are, however, also worth considering for ontology design, though we won't call out specific examples here.

SCENARIOS

There are a couple of key scenarios to consider.

1. You are designing a form in a single language (let's assume English) that people from around the world will be filling in.
2. You are designing a form in one language but the form will be adapted to suit the cultural differences of a given locale when the site is translated.

In reality, you will probably not be able to localize for every different culture, so even if you rely on approach 2, some people will still use a form that is not intended specifically for their culture.

EXAMPLES OF DIFFERENCES

To get started, let's look at some examples of how people's names can be different around the world.

Given name and patronymic

In the Icelandic name **Björk Guðmundsdóttir** Björk is the given name. The second part of the name indicates the father's (or sometimes the mother's) name, followed by -son for a male and -dóttir for a female, and is more of a description than a family name in the Western sense. Björk's father, Guðmundur, was the son of Gunnar, so is known as Guðmundur Gunnarsson.

Icelanders prefer to be called by their given name (Björk), or by their full name (Björk Guðmundsdóttir). Björk wouldn't normally expect to be called Ms. Guðmundsdóttir. Telephone directories in Iceland are sorted by given name.

Other cultures where a person has one given name followed by a patronymic include parts of Southern India, Malaysia and Indonesia.

In the Malay name **Isa bin Osman** the word 'bin' means 'son of' ('binti' is used for women). If you refer to this person you might say Mr. Isa, or if you know him personally, Encik Isa (Encik is a Malay word rather like Mr.).

Different order of parts

In the Chinese name **毛泽东** (Mao Ze Dong) the family name is Mao, ie. the *first* name when reading (left to right). The given name is Dong. The middle character, Ze, is a generational name, and is common to all his siblings (such as his brothers and sister, 毛泽民 (Mao Ze Min), 毛泽覃 (Mao Ze Tan), and 毛泽红 (Mao Ze Hong)).

Among people who are not on familiar terms, Mao may be referred to as 毛泽东先生 (Mao Ze Dong xiān shēng) or 毛先生 (Mao xiān shēng) (xiān shēng being the equivalent of Mr.). Although not everyone has a generational name these days, especially in Mainland China, those who do have one expect it to be used together with their given name. Thus, if you are on familiar terms with someone called 毛泽东, you would normally refer to them using 泽东 (Ze Dong), not just 东 (Dong).

Note also that the names are not separated by spaces.

Other cultures, such as in Japan, Korea, and Hungary, also order names as family name followed by given name(s).

Chinese people who deal with Westerners will often adopt an additional given name that is easier for Westerners to use. For example, Yao Ming (family name Yao, given name Ming) may write his name for foreigners as Fred Yao Ming or Fred Ming Yao.

Multiple family names

Spanish-speaking people will commonly have two family names. For example, **María José Carreño Quiñones** (José being a part of her given name) may be the daughter of Antonio Carreño Rodríguez and María Quiñones Marqués.

You could refer to her as Señorita Carreño, but not Señorita Quiñones. However, more recently there is also a preference to move away from titles that indicate the married status of women (especially when they are not so indicated for men), so 'Señorita' may be replaced with 'Señora'.

Brazilians have similar customs, and may even have three or four family names, drawing on the names of other ancestors, such as **José Eduardo Santos Tavares Melo Silva**.

Typically, two Spanish family names would have the order paternal+maternal, whereas Portuguese names in Brazil would be maternal+paternal. However, this order may change.

Furthermore, some names add short words, such as de or e between family names, such as Carreño de Quiñones, or Tavares e Silva.

Variant word forms

We already saw that the patronymic in Iceland ends in -son or -dóttir, depending on whether the child is male or female. Russians use patronymics as their middle name but also use family names, in the order givenName-patronymic-familyName. The endings of the patronymic and family names will indicate whether the person in question is male or female. For example, the wife of **Борис Николаевич Ельцин** (Boris Nikolayevich Yeltsin) is **Наина Иосифовна Ельцина** (Naina Iosifovna Yeltsina) – note how the husband's names end in consonants, while the wife's names (even the patronymic from her father) end in -a.

By the way, a slightly less formal way of writing Russian names follows the order familyName-givenName-patronymic, such as **Ельцина Наина Иосифовна**.

Middle initials

Americans often write their name with a middle initial, for example, **John Q. Public**. Often forms designed in the USA assume that this is common practice, whereas even in the UK, where people may indeed have (one or more) middle names, this is often seen as a very American approach. People in other countries who have more than two names and don't usually initialise them may be confused about how to deal with such forms. Bear in mind, also, that many people who do use an initial in their name may use it at the beginning.

Filipinos also write their name with a middle initial, but it represents the mother's name before marriage rather than a given name. For example, in **Maria J. Go**, the initial represents Jimenez, the previous family name of Maria's mother. (In fact, an initial may represent more than one name: 'D' may stand for 'Dela Cruz' when the name is written in full.)

Inheritance of names

It would be wrong to assume that members of the same family share the same family name. There is a growing trend in the West for wives to keep their own name after marriage, but there are other cultures, such as China, where this is the normal approach. In some countries the wife may or may not take the husband's name. If the Malay girl Zaiton married Isa, mentioned above, she may remain Mrs. Zaiton, or she may choose to become Zaiton Isa, in which case you might refer to her as Mrs. Isa.

Some Hispanic names approach this slightly differently. In 1996 Manuel A. Pérez Quiñones described [1] the names in his family. As mentioned above, his family names, known as *apellidos*, became Pérez Quiñones because his father's *apellidos* were Pérez Rodríguez and his mother's *apellidos* were Quiñones Alamo. In time, he courted a girl with the *apellidos* Padilla Falto. When they married, her *apellidos* became Padilla de Pérez. Their children were called Pérez Padilla, and so on. The point here is that only the children in the family have the same *apellidos*.

i I have slightly changed Manuel's name here. In fact, he typically spells his name with a hyphen (Pérez-Quíñones) to get around confusion in the USA about how his name works. Others may do this too.

You should also not simply assume that name adoption goes from husband to wife. Sometimes men take their wife's name on marriage. It may be better, in these cases, for a form to say 'Previous name' than 'Maiden name' or 'née'.

Mixing it up

Many cultures mix and match these differences in personal names, and add their own novelties.

For example, **Velikkakathu Sankaran Achuthanandan** is a Kerala name from Southern India, usually written **V. S. Achuthanandan** which follows the order `familyName-fathersName-givenName`.

In many parts of the world, parts of names are derived from titles, locations, genealogical information, caste, religious references, and so on. Here are a few examples:

- the Indian name **Kogaddu Birappa Timappa Nair** follows the order villageName-fathersName-givenName-lastName.
- the Rajasthani name **Aditya Pratap Singh Chauhan** is composed of givenName-fathersName-surname-casteName.
- in another part of India the name **Madurai Mani Iyer** represents townName-givenName-casteName.
- the Arabic **Abu Karim Muhammad al-Jamil ibn Nidal ibn Abdulaziz al-Filistini** translates as "Father of Karim, Muhammad (given name), The beautiful, Son of Nidal, Son of Abdulaziz, the Palestinian". Karim is Muhammad's first-born son. (For more details about this rich naming tradition, see Wikipedia [2].)

In Thailand people have a nickname, that is usually not related to their actual name, and will generally use this name to address each other in non-formal situations. (They will also typically introduce themselves to Westerners with this name, since it is usually only one or two syllables and therefore easier to pronounce.) Former prime minister **Thaksin Shinawatra** has the nickname **Maew** (แมว). Often they will have different nicknames for family and friends.

In Vietnam, names such as **Nguyễn Tấn Dũng** follow the order familyName-middleName-givenName. Although this seems similar to the Chinese example above, even in a formal situation this Prime Minister of Vietnam is referred to using his *given name*, ie. Mr. Dũng, not Mr. Nguyễn.

Ambiguity in written forms

Ideographic characters in Japanese names can typically be pronounced in more than one way. In some cases this makes it difficult for people to know exactly how to pronounce a name, and also causes problems for automatic sorting and retrieval of names, which is typically done on the basis of how the name is pronounced. For example, the family name of **東海林賢蔵** (ie. the first three ideographic characters on the left) may be transcribed or pronounced as either Tōkairin or Shōji.

Furthermore, different kanji characters may be pronounced in the same way, so romanization (ie. Latin script transcription) tends to lose important distinctive information related to names. For example, 庄司, 庄子, 東海林, and 小路 can all be romanized as Shōji.

Some Japanese names use archaic ideographic characters, or characters that are no longer used in modern Japanese. The pronunciation of these characters may not be recognized.

Because of these issues, Japanese people will commonly provide a phonetic version of their name (using a non-ideographic Japanese kana alphabet) along with the normal written version.

Further information

The information above uses only simple cases to describe a number of significant divergences in the way people construct names. The reality, even within a single culture, is typically even more complicated. Wikipedia sports a large number of fascinating articles about how people's names look in various cultures around the world. It is recommended that you read more detailed the information accessed via the following links.

Akan [3] • Arabic [4] • Balinese [5] • Bulgarian [6] • Czech [7] • Chinese [8] • Dutch [9] • Fijian [10] • French [11] • German [12] • Hawaiian [13] • Hebrew [14] • Hungarian [15] • Icelandic [16] • Indian [17] • Indonesian [18] • Irish [19] • Italian [20] • Japanese [21] • Javanese [22] • Korean [23] • Lithuanian [24] • Malaysian [25] • Mongolian [26] • Persian [27] • Philippine [28] • Polish [29] • Portuguese [30] • Russian [31] • Spanish [32] • Taiwanese [33] • Thai [34] • Vietnamese [35]

IMPLICATIONS FOR FIELD DESIGN

As mentioned above, one possible approach is to localize forms for a particular culture. In theory this should allow you to tailor your forms exactly

to the needs of the audience. Unfortunately, there may still be a number of possible disadvantages to this approach:

- If you need to centralise data from several locales within a single database, using localized form layouts will simply defer the difficulties of synthesizing the information across cultures until the time when you need to store the data.
- Even within a single country people will typically have different ways of forming personal names. For example there may be foreigners living in the country, there may be different cultural elements within the country (eg. Singaporeans have names of Chinese, Malay and South Indian origin), or there may just be more than one way of using names. Therefore your forms will often need to allow for some flexibility.

In what follows we propose some general guidelines that may help. Unfortunately, this is a complex topic and the suggestions here are for the very general case, and don't address all the issues.

To split or not to split?

If designing a form or database that will accept names from people with a variety of backgrounds, you should **ask yourself whether you really need to have separate fields for given name and family name** .

This will depend on what you need to do with the data, but obviously it will be simpler, where it is possible, to just use the full name as the user provides it.

YOUR PROFILE

Full name

Bear in mind that names in some cultures can be quite a lot longer than your own. **Make input fields long enough to enter long names, and ensure that if the name is displayed on a web page later there is enough space for it** . Also **avoid limiting the field size for names in your database**. In particular, do not assume that a four-character Japanese name in UTF-8 will fit in four bytes – you are likely to actually need 12.

Strategies for splitting up names

If you do still feel you need to ask for constituent parts of a name separately, **try to avoid using the labels 'first name' and 'last name' in non-localized forms** , since these can be confusing for people who normally write their family name followed by given names.

YOUR PROFILE

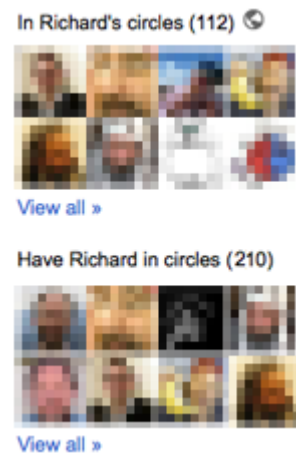
Family name

Other/given names

For some cultures this is still problematic (for example Icelanders, who don't actually have family names), but, short of very localized customization, this is probably the best we can make a generic form.

In some cases you want to identify parts of a name so that you can sort a list of names alphabetically, contact them, etc. **Consider whether it would make sense to have one or more extra fields, in addition to the full name field, where you ask the user to enter the part(s) of their name that you need to use for a specific purpose.**

Sometimes you may opt for separate fields because you want to be able to use part of the name to address the person directly, or refer to them. For example, when Google+ refers to "Richard's contacts". Or perhaps it's because you want to send them emails with their name at the top. Note that not only may you have problems due to name syntax here, but you also have to account for varying expectations around the world with regards to formality (not everyone is happy for a stranger to call them by their given name). It may be better to **ask separately, when setting up a profile for example, how that person would like you to address them** .



YOUR PROFILE

Full name

What should we call you? (for example, when we send you mail?)

This extra field would also be useful for finding the appropriate name from a long list, and for handling Thai nicknames.

By the way, for sorting Japanese names you will need an additional field for them to type how their name is pronounced, since you can't always tell how to pronounce it from the ideographic characters. Such pronunciation information is used for sorting Japanese names.

Also, **if you have separate fields for parts of a person's name, ensure that you label clearly which parts you want where.** For example, don't assume that the order they will provide names in will be given followed by family.

Be careful, also, about assumptions built into algorithms that pull out the parts of a name automatically . For example, the v-card and h-card approach of implied "n" optimization [36] could have difficulties with, say, Chinese names. You should **be as clear as possible about telling people how to specify their name** so that you capture the data you think you need.

Don't assume that a single letter name is an initial. People do have names that are one letter long. These people can have problems if the form validation refuses to accept their name and demands that they supply their name in full. If you want to encourage people not to use initials, perhaps you should make that a warning message, rather than block the form submission.

Similarly, **don't require that people supply a family name.** In cultures such as parts of Southern India, Malaysia and Indonesia, a large number of people have names that consist of a given name only, with no patronym. If you require family names, you may create significant problems in these cultures, as users enter garbage data like "." or "Mr." in the family name field just to escape the form.

Other things

Don't forget to allow people to use punctuation such as hyphens, apostrophes, etc. in names. Don't require names to be entered all in upper case – this can be difficult on a mobile device. **Allow the user to enter a name with spaces** , eg. to support prefixes and suffixes such as de in French, von in German, and Jnr/Jr in American names, and also because some people consider a space-separated sequence of characters to be a single name, eg. Rose Marie.

Don't normalize the casing in names. Some names (such as 'McNamara') contain capital letters that are not the first letter; others (such as 'van der Waals') include words that are not capitalized. Forms should preserve the case the user enters and not coerce such names to always and only use capital letters at the start of each word.

Don't assume that members of the same family will share the same family name.

Ask yourself whether it's necessary to require that people supply a title. Some people don't expect to use titles at all. In that case, they should be able to leave such a field blank. If you are requiring people to supply a title such as Mr./Mrs./Ms. in order to indicate their gender, why not ask directly? This also gets around the issue that you may be inadvertently asking women to take decisions about revealing their marital status, but not men. Of course, also bear in mind that titles vary significantly around the world (see below).

As mentioned earlier, because it is not only women who change their family names, **it may be better for a form to ask for 'Previous name' rather than 'Maiden name' or 'née'** .

If you are designing forms that will be localized on a per culture basis, don't forget that atomized name parts may still need to be stored in a central database, which therefore needs to be able to represent all the various complexities that you dealt with by relegating the form design to the localization effort.

IMPLICATIONS FOR CHARACTER SUPPORT

The first thing that English speakers must remember about other people's names is that a large majority of them don't use the Latin alphabet, and of those that *do*, a majority use accents and characters that don't occur in English. It seems obvious, once it is said, but it has some important consequences for designers that are often overlooked.

If you are designing an English form you need to decide whether you are expecting people to enter names in their own script (eg. 小林康宏) or in an Latin-only transcription (such as Yasuhiro Kobayashi), or both.

Remember that even names in English may involve non-ASCII characters (eg. Zoë).

On the other hand, there are some situations, such as a login name on an ASCII-only system, where you can't permit non-ASCII characters.

What people will type into the form will often depend on whether the form and its page is in their language or not. If the page is in their language, don't be surprised to get back non-Latin or accented Latin characters.

i In terms of letters, ASCII-only means using the basic letters of the English alphabet, ie. ABCDEFGHIJKLMNOPQRSTUVWXYZ (upper- and lowercase).

If you hope to get Latin- or ASCII-only, you need to tell the user.

! Don't forget to tell the people who are translating your page to explain this to users.

YOUR PROFILE

Full name (Use only letters A to Z with no accents.)

The decision about which is most appropriate will depend to some extent on what you are collecting people's names for, and how you intend to use them.

- Are you collecting the person's name just to have an identifier in your system? If so, it may not matter whether the name is stored in ASCII-only or native script.
- Or do you plan to call them by name on a welcome page or in correspondence? If you will correspond using their name on pages written in their language, it would seem sensible to have the name in the native script.
- Is it important for people in your organization who handle queries to be able to recognise and use the person's name? If so, you may want to ask for a Latin transcription.
- Will their name be displayed or searchable (for example Flickr optionally shows people's names as well as their user name on their profile page)? Or will you want to send them correspondence in their own language, but track them in your back-office in a language such as English? If so, **you may want to store the name in both Latin and native scripts, in which case you probably need to ask the user to submit their name in both native script and Latin-only form, using separate fields** .

YOUR PROFILE

Name (in your alphabet)

Latin transcription (if different)

Note that Japanese users may need to provide a transcription in a Japanese syllabic script rather than/in addition to the ideographic form. This could lead to a third field in the example above.

If you do accept non-ASCII names, you should use a Unicode character encoding (eg. UTF-8) in your pages, your back end databases and in all the software code in between. This will significantly simplify your life.

BY THE WAY

A note on sorting

Lists of names are not always sorted by family name around the world. For example, Thai and Icelandic people expect lists to be sorted by given name instead.

In another example, it is possible that sort orders can also be different in different parts of the Spanish-speaking world. For instance, María-Jose Carreño Quiñones in one place would expect to find her name in a list by looking up Carreño Quiñones. María-Jose Carreño Quiñones from another place, however, would expect her name to be sorted by Quiñones.

The treatment of small words such as "von", "de", and "van" brings additional complexity to sorting. Sometimes the prefixes are significant, other times they are not.

Formality and honorifics

Different levels of formality apply in different cultures. When addressing someone you need to take this into account. Whereas given names are becoming a popular form of address in Western and technology circles, it is by no means universally appropriate. Contacting someone for the first time in the UK using their given name can sometimes imply that you have previously met them.

On the other hand, addressing someone using a title and given name (eg. "Mr. Edward") or just by their family name (eg. "Windsor!") are acceptable in some parts of the world, but not in others (such as the UK).

In Germany, titles are important, and you may need to refer to someone as not just Herr Schmidt, but Herr Professor Doktor Schmidt.

In a culture such as that in Japan, it is normal to add an honorific or job title to the name of someone you contact. For example, it would be expected to refer to someone as Tanaka-san or Tanaka-sama (depending on your relationship to them). A departmental manager named Tanaka would expect to be referred to as Tanaka-bucho (Department-head Tanaka). Although you can attach -san to given names, it would be very unusual to do so with people in the work environment.