


What's In A Name?

Addressing people by the right name around the world

Kip Cole



Informal Surname
Given Name

Attribute	Value
Name Locale	en-AU
Name order	given_first
Formality	informal
Usage	addressing
Format:	medium

CLDR-Based

- [ex_cldr](#) is the base level library that manages language tags and the locale data that supports them. Over 500 locales are supported.
- [ex_cldr_numbers](#) which provides localized number formatting and parsing.
- [ex_cldr_currencies](#) which provides the data about the world's currencies both current and historic.
- [ex_cldr_dates_times](#) which provides localized date and time formatting (but not parsing)
- [ex_cldr_units](#) which provides localized units-of-measure formatting and parsing.
- [ex_cldr_units_sql](#) which provides a database type and an Ecto type for serializing units-of-measure.
- [ex_cldr_messages](#) implements the Unicode Message format for localizing messages. Integrates with Gettext.
- [ex_cldr_calendars](#) which provides localized calendar implementations of the Proleptic Gregorian calendar, the Julian calendar, the ISO Week calendar and provides a mechanism to define variations of the Gregorian and ISO Week calendars to meet the needs to organizations (like corporations or governments) that use different calendar periods.
- [ex_cldr_lists](#) implemented localized list formatting.
- [ex_cldr_territories](#) by @Schultzer provides support for localizing territory (country) information.
- [ex_cldr_languages](#) by @LostKobrakai which provides support for localizing language names.
- [ex_cldr_collation](#) which implements the Unicode Collation Algorithm although only the default DUCET collation.

<https://github.com/elixir-cldr>

<https://github.com/kipcole9/money>

Money

- [ex_money](#) provides a full-feature money libraries focusing on correctness and reliability with support for currency specific formatting (in any locale), money parsing, amortization and basic math.
- [ex_money_sql](#) provide a database type and an ecto type to simplify serialization of money without losing precision or currency tags. Also includes some aggregate function extensions for the Postgresql database.

Ecto Integration for database translations

- [ex_cldr_trans](#) provides a mechanism to localized database content. It is based on the fabulous [trans](#) by @crbelaus.

Plug-based application support

- [ex_cldr_plugs](#) provides plugs that can extract the requested locale for a user from different parts of the locale or the session.
- [ex_cldr_routes](#) provides functions to generate and recognize localized routes. Also provides localized route helpers. Does not yet support localized [verified routes](#). Verified routes are expected to be supported in March 2023.
- [ex_cldr_html](#) which provides form helpers for selecting currencies, languages and territories. These are static HTML generators. A future `ex_cldr_components` will provide a more complete LiveView localized UI experience. This is not expected before year end 2023.

Unicode

- [unicode](#) which provides introspection into the Unicode database
- [unicode_set](#) which provides set operations on unicode attributes such as categories, blocks, scripts, It includes a regex engine that supports the Unicode regex standard (mostly).
- [unicode_string](#) which locale-aware case folding, case mapping (upcase, downcase and titlecase) case-insensitive equality as well as word, line, grapheme and sentence breaking and streaming.
- [unicode_guards](#) Implements Unicode Set-based guards for Elixir. Supports matching unicode sets to codepoints that can be used in function guards..
- [unicode_unihan](#) functions to introspect the Unicode Unihan character database. By @jwchui.

Image

- Image is an approachable image processing library primarily based upon Vix and libvips that is NIF-based, fast, multi-threaded, pipelined and has a low memory footprint.

Astro

- Astronomical calculations in Elixir including sunrise, sunset, moonrise, moonset, equinox, solstice, moon phase and more.

Tz_world

- Resolve time zone names from a geographic location

Kday

- Calculations for the first, last or nth day of the week on, nearest, after or before a date.


<https://github.com/elixir-cldr>
<https://github.com/elixir-unicode>
<https://github.com/elixir-image>
<https://github.com/kipcole9>

Adam



Name

Adam & Eve



Name Name

盤古

Name

<https://en.wikipedia.org/wiki/Pangu>

मनु

Name

[https://en.wikipedia.org/wiki/Manu_\(Hinduism\)](https://en.wikipedia.org/wiki/Manu_(Hinduism))

Meanwhile, in the middle ages in England

John
Name
7.7%

William
Name
14%

John the butcher



Given Name



Byname

Cognominal

Based upon appearance,
temperament

- Schwartzkopf
- Short
- Caesar

Occupational

Based upon occupation

- Butcher
- Baker
- Miller

Topynomic

Based upon appearance,
temperament

- Schwartzkopf
- Short
- Caesar

Ornamental

Based upon sounding
like a noble

- Gyldenstjerne
("golden star")
- Rosenkrantz ("rose
wreath")

Patronymic & Matronymic

Based upon parents' names

- | | |
|-----------------|-----------------|
| • Andersen | • Stefanović |
| • O'Brien | • Petrov |
| • Ali Mohamed | • Dēmētrópoulos |
| • Stefánsdóttir | • Fitzgerald |

Charles Emerson Winchester

First Name

Middle Name

Last Name

Charles Emerson Winchester III

First Name

Middle Name

Last Name

Generation

Charles Emerson Winchester III M.D.

First Name

Middle Name

Last Name

Gen.

Credentials

Chuck

Informal First Name

A nickname of Charles, which comes from the Germanic word *cheorl*, meaning "free man". Could also be from the Old English *chukken*, meaning "to cluck".

<https://www.babycentre.co.uk/babynome/1023371/chuck#>

Sr. Miguel Ángel Juan Antonio Pablo Perez II



Title Given Name Other Given Names Family Name Other Family Names Generation

Johannes Adam Ferdinand Alois Josef Maria Marko d'Aviano Pius
von und zu Liechtenstein.

Surname prefix
Tussenvoegsel

[https://en.wikipedia.org/wiki/Hans-Adam II, Prince of Liechtenstein](https://en.wikipedia.org/wiki/Hans-Adam_II,_Prince_of_Liechtenstein)

<https://en.wikipedia.org/wiki/Tussenvoegsel>

พระบาทสมเด็จพระเจ้าอยู่หัวภูมิพลอดุลยเดช

Phra Bat Somdet Phra Chao Yu Hua Phumiphon Adunyadet

Title

https://en.wikipedia.org/wiki/Thai_royal_ranks_and_titles

Different way to construct names

- There is a wide variety in the way that people's names appear in different languages.
- People may have a different number of names, depending on their culture—they might have only one name (“Zendaya”), two (“Albert Einstein”), or three or more.
- People may have multiple words in a particular name field, eg “Mary Beth” as a given name, or “van Berg” as a surname.
- Some languages, such as Spanish, have two surnames (where each can be composed of multiple words).
- The ordering of name fields can be different across languages, as well as the spacing (or lack thereof) and punctuation.
- Name formatting needs to be adapted to different circumstances, such as a need to be presented shorter or longer; formal or informal context; or when talking about someone, or talking to someone, or as a monogram (JFK).

Unicode CLDR Project



News

- [2024-03-05](#) [CLDR v45 alpha available for testing](#)
- [2023-12-13](#) [CLDR v44.1 released \(an update to CLDR v44\)](#)
- [2023-10-31](#) [CLDR v44 released](#)

What is CLDR?

The Unicode Common Locale Data Repository (CLDR) provides key building blocks for software to support the world's languages, with the largest and most extensive standard repository of locale data available. This data is used by a [wide spectrum of companies](#) for their software internationalization and localization, adapting software to the conventions of different languages for such common software tasks. It includes:

- **Locale-specific patterns for formatting and parsing:** dates, times, timezones, numbers and currency values, measurement units,...
- **Translations of names:** languages, scripts, countries and regions, currencies, eras, months, weekdays, day periods, time zones, cities, and time units, emoji characters and sequences (and search keywords),...
- **Language & script information:** characters used; plural cases; gender of lists; capitalization; rules for sorting & searching; writing direction; transliteration rules; rules for spelling out numbers; rules for segmenting text into graphemes, words, and sentences; keyboard layouts...
- **Country information:** language usage, currency information, calendar preference, week conventions,...
- **Validity:** Definitions, aliases, and validity information for Unicode locales, languages, scripts, regions, and extensions,...

[News](#)

[What is CLDR?](#)

[Who uses CLDR?](#)

[How to Use?](#)

[How to Contribute?](#)

[Translations and other language data](#)

[Code and Structure](#)

[Tickets](#)

[Who has contributed?](#)

[What is the Schedule?](#)

<https://cldr.unicode.org>

Structure of a name

```
iex> t Cldr.PersonName.t
@type t() :: %Cldr.PersonName{
  credentials: String.t() | nil,
  generation: String.t() | nil,
  given_name: String.t() | nil,
  informal_given_name: String.t() | nil,
  locale: Cldr.Locale.locale_reference(),
  other_given_names: String.t() | nil,
  other_surnames: String.t() | nil,
  preferred_order: Cldr.PersonName.Formatter.name_order() | nil,
  surname: String.t() | nil,
  surname_prefix: String.t() | nil,
  title: String.t() | nil
}
```

Three ways to provide data to the formatter:

- `Cldr.PersonName.t` struct
- `Cldr.PersonName` behaviour
- `Cldr.PersonName.Format` protocol

How we use names

- Formality, usage, name order, ...

Formality

- *Formality* indicates the formality of usage. A name on a badge for an informal gathering may be much different from an award announcement at the Nobel Prize Ceremonies.
- Note that the formats may be the same for different formality scenarios depending on the length, usage, and cultural conventions for the locale. For example, short formal and short informal may both be just the given name.

Formality	Description
formal	<p>A more formal name for the individual. The composition depends upon the language. For example, a particular locale might include the title, generation, credentials and a full middle name (given2) in the long form.</p> <p>length="medium", formality="formal" "Robert J. Smith"</p>
informal	<p>A less formal name for the individual. The composition depends upon the language. For example, a language might exclude the title, credentials and given2 (middle) name. Depending on the length, it may also exclude the surname. The formatting algorithm should choose any passed in name data that has an <i>informal</i> attribute, if available.</p> <p>length="medium", formality="informal" "Bob Smith"</p>

Usage

The usage indicates if the formatted name is being used to address someone, refer to someone, or present their name in an abbreviated form.

Usage	Description
addressing	Used when speaking “to” a person, or “vocative” case. This may also have an effect on the formality. example: “Welcome, Robert ”
referring	Used when speaking “about” a person, or “nominative” case. example: “ Robert Smith joined your group”
monogram	The monogram usage is for a specific abbreviated form for computer UI. Example: a monogram for Robert James Smith may be RS or RJS .

Name Order

- The `order` attribute is used for patterns with different orders of fields.
- The *sorting* patterns are chosen based on input parameters, while the choice between *given_first* and *surname_first* is based on features of the **PersonName** object to be formatted and the *name_order* element values for a locale.
- For example, when the display language is Japanese, it is customary to use *surname_first* for names of people from Japan and Hungary, but use *given_first* for names of people from the United States and France. Although the English pattern for sorting is distinct from the other patterns (except for unusual names), that is not necessarily the case in other languages.

Name Order	Description
given_first	The given name precedes the surname.
surname_first	The surname precedes the given name.
sorting	Used to format names for a sorted list. example: “Brown, William” [medium, informal]

Formatting a name

- **The formatting locale.** This is used to choose the primary set of patterns to format name data.
- **The name locale.** If the name data comes from a locale different from the formatting locale, it may need to be handled differently. If the name locale is not known, an inferred name locale is derived from the information in the name and the formatting locale.
- **Input parameters.**
 - ***order*** - indicates whether the given name comes first or the surname. This is normally specified in the CLDR data for the locale. This feature is also used for the sorting format.
 - ***format*** - used to select patterns for common short, medium, and long formatted names.
 - ***usage*** - this is used to select the correct pattern to format a name when a program is *addressing* or talking **to** a person or it is *referring* to or talking **about** another person.
 - ***formality*** - This is used to select the formal or informal formatting of a name.

Format templates in CLDR

Example from the "es" locale

```
<personName order="sorting" length="long" usage="referring" formality="formal">
  <namePattern>{surname}, {title} {given} {given2}</namePattern>
  <namePattern alt="1">{surname} {surname2}, {title} {given} {given2}</namePattern>
</personName>
<personName order="sorting" length="long" usage="referring" formality="informal">
  <namePattern>{surname} {surname2}, {given-informal}</namePattern>
  <namePattern alt="1">{surname} {surname2}, {given-informal}</namePattern>
</personName>
<personName order="sorting" length="medium" usage="referring" formality="formal">
  <namePattern>{surname}, {title} {given} {given2-initial}</namePattern>
  <namePattern alt="1">{surname} {surname2}, {title} {given} {given2-initial}</namePattern>
</personName>
<personName order="sorting" length="medium" usage="referring" formality="informal">
  <namePattern>{surname}, {given-informal}</namePattern>
  <namePattern alt="1">{surname} {surname2}, {given-informal}</namePattern>
</personName>
```

Down the rabbit hole

- Initialisation of a name requires knowing how to break words
 - That requires Unicode break algorithm -> [unicode_strings](#)
 - Which needs Unicode sets -> [unicode_sets](#)
 - Which needs enhanced regular expressions -> [unicode_regex](#)
 - Which needs Unicode introspection -> [unicode](#)

Livebook examples

https://github.com/elixir-cldr/cldr_person_names/blob/main/person_name_formatting_explorer.livemd

What's Not in CLDR scope

- Grammatical inflection of formatted names.
- Context-specific cultural aspects, such as when to use “-san” vs “-sama” when addressing a Japanese person.
- Providing locale-specific lists of titles, generation terms, and credentials for use in pull-down menus or validation (Mr, Ms., Mx., Dr., Jr., M.D., etc.).
- Validation of input, such as which fields are required, and what characters are allowed.
- Combining alternative names, such as multicultural names in Hong Kong "[Jackie Chan Kong-Sang](#)", or 'Dwayne “The Rock” Johnson' [However, *personas*]
- More than two levels of formality for names.
- Parsing of names.

Parsing is hard

Parsing of name strings into specific name parts such as *given_name* and *other_given_names* is ambiguous. A name like "Mary Beth Estrella" could conceivably be any of the following.

Given Name	Other Given Names	Surname	Other Surnames
Mary	Beth	Estrella	
Mary Beth		Estrella	
Mary		Beth Estrella	
Mary		Beth	Estrella

What's not covered

But I'd like to find a way

Salutations

Illustrissime et Éminentissime Seigneur

(Most Illustrious and Most Eminent Lord)

Valedictions

Veillez agréer, Madame, Monsieur, l'expression de mes sentiments distingués"

(Please accept, Madam, Sir, the expression of my distinguished sentiments)

Personas

<https://en.wikipedia.org/wiki/Salutation>

<https://en.wikipedia.org/wiki/Valediction>

Timetable

- CLDR 45 will be released in April. See <https://cldr.unicode.org/index/downloads/cldr-45>
- ex_cldr 2.38 and all necessary updates will be out the same day as CLDR 45 is released
- ex_person_names will be available on the same day
 - Available now for experimentation by configuring as:
`{:ex_cldr_person_names, github: "elixir-cldr/cldr_person_names"}`



```
contacts = %{  
  elixir_forum: "@kip",  
  elixir_slack: "@kip",  
  github: "kipcole9"  
}
```