Appendices

Appendix A – Annotation Guideline

Tag sets

First_Name

An individual's first name. Only one tag should be assigned to each individual's first name, i.e., double names should be included in one tag.

<First_Name>Per Olav</First_Name>

Last_Name

An individual's last name, including middle names that could be considered a last name. Any academic title (such as Dr.) should not be included in the annotation. Even double last names should be included in one tag.

Dr <Last_Name>Ebbesen Dyrdal</Last_Name>

Age

Any age written with numbers, letters, or a mixture of both.

<Age>46 år</Age>

Hun ble <Age>fire måneder</Age> gammel.

<Age>100-åringen</Age>

Health_Care_Unit

All healthcare institutions where health care is provided (hospitals, healthcare centers, nursing homes, etc.) and their clinical units (emergency department, neuroclinic, department of medical genetics, etc.). Even abbreviations and other unofficial institution names should be included inside this tag. Annotate with a largest possible span (as long as each word inside the annotation should be a part of the annotation).

<Health_Care_Unit>avdeling for medisinsk genetikk, Ullevål</Health_Care_Unit>

<Health_Care_Unit>avdeling for medisinsk genetikk</Health_Care_Unit> på

<Health_Care_Unit>Ullevål</Health_Care_Unit>

Social_Security_Number

A person's social security number.

<Social_Security_Number>04126150397</Social_Security_Number>

Phone_Number

A phone number. Any mentioned country codes should also be included inside the tag.

<Phone_Number>+47 90491855</Phone_Number>

Date_Full

Including the combination of date, month, and year, written with numbers, letters, or a mixture of both.

<Date_Full>01. mai 1979</Date_Full>

Date_Part

Including one or two of the instances date, month, and year, written with numbers, letters, or a mixture of both. Does not include weekdays, week numbers, or seasons.

<Date_Part>januar 2008</Date_Part>

Location

Any geographical location, including countries, cities, towns, post codes, addresses, etc. Separate different locations into different tags.

Han bor i <Location>Stockholm</Location> <Location>Sverige</Location>.

<Location>Hovseterveien 26</Location>, <Location>0157 Oslo</Location>

<Health_Care_Unit>Østfold sykehus</Health_Care_Unit> i <Location>Moss</Location>

<Location>Hovseterveien 24B</Location>, <Location>2940 Oslo</Location>

General annotation instructions

- It is perfectly fine to include more than one token in one annotation, but do not split a token. A token is defined as a string of characters between two spaces or symbols. However, full stops and hyphens inside a word, and percentage symbols located to the right of a number, are defined as part of the word which makes up one token. Symbols next to a word and another symbol or space, on the other hand, such as full stops, constitute their own tokens. To visualize, the sentence "Han (70-åringen) fikk slag 11.11.12." consists of the following tokens (separated with line breaks):

Han

(

70-åringen

)

fikk

slag

11.11.12

.

Correct: *<Age>30-års</Age>*

Not correct: *<Age>30</Age>-års*

- A token can only be assigned to one tag. Choose the one that you think is correct for the specific sentence. Ask yourself from what category you could find a pseudonym.

Correct: *<First_Name>Lars</First_Name> <Last_Name>Holmen</Last_Name>*

Not correct: *<First_Name>Lars</First_Name> <Location>Holmen</Location>*