

# On Time Coverage in the Virtual Language Observatory

Jörg Knappen\*

2014-05-14

## 1 Summary

Time coverage can be represented by a *named period*, a *date*, or a *date range*. In this note we ignore the named period approach. Dates are represented by the according XSL data types *xsd:date*, *xsd:dateTime*, and *xsd:time*. Date ranges can be represented by a component using begin and end dates, or directly by the *Open Date Range Format (ODRF)*.

Relevant data categories for time coverage are DC-2502, -3664, -4343. For begin and end of time intervals, the data categories DC-3654 and -3655 are pertinent. Time coverage with begin and end time points is implemented by the component c\_1349361150714.

## 2 On time coverage

The time coverage is defined as “The [...] temporal topic of the resource.”<sup>1</sup> It is further explained as “Temporal topic may be a named period, date, or date range.”

In this note we will not deal with *named period* as a designator of time coverage because it is hard to formalise and to check automatically. The data range indicated by a named period (like renaissance) may have fuzzy boundaries that are hard to interpret by a computer. So there are *date* and *date range* to deal with. Date ranges may be *closed* having a begin and an end or *open* having a start and no end (after some date) or having an end but no start (before some date).

A certain resource, e.g., a diachronic corpus may cover several discrete date ranges.

Note that a date range is different from a *duration* which has a fixed length but no fixed start or end dates.

---

\*CLARIN VLO Task Force

<sup>1</sup><http://dublincore.org/documents/dcmi-terms/#terms-coverage>

### 3 Use cases

A user may look for

- Ressources in middle polish from the 17th century
- Recent resources in german language
- Ressources from the French revolution and the Napoleonic era

The last example illustrates that a faceted approach (starting with centuries, going down to decades and years) might not be the best choice for a user interface.

### 4 Proposed user interface

A simple user interface can consist of two input fields, one for the begin of temporal coverage and one for the end of temporal coverage.

Leaving one of the input fields empty means that the date range is open at this end.

In addition, there is a radio button with values “contains” and “overlaps”. The choice “contains” means that the coverage of the resource lies completely in the user’s selection. This maximises Precision at the cost of Recall. The choice “overlaps” means that the coverage of the resource has some overlap with the user’s selection, but may be larger. This maximises Recall at the cost of Precision.

A graphical approach to a user interface can be a display of the time axis with two sliders to select begin and end of a date range.

## 5 Vocabularies

### 5.1 Date

For the date there are *xsd:date*, *xsd:time*, and *xsd:dateTime* defined by the w3c consortium based on the international standard ISO 8601.<sup>2</sup>

Note that the XML specification is not exactly identical to ISO 8601. Most importantly, there is no year 0000 in the XML spec. The year 1 BC is -0001 in the XML spec.

### 5.2 Date Range

For date ranges there is the *Open Date Range Format (ODRF)*<sup>3</sup>.

The ODRF can describe both closed and open date ranges.

---

<sup>2</sup><http://www.w3.org/TR/xmlschema-2/#isoformats>

<sup>3</sup><http://www.ukoln.ac.uk/metadata/dcmi/date-dccd-odrf/>

The ODRF is different from the description of *time intervals* in ISO 8601 in two respects: (i) It restricts the notation to the simplest possible cases (ii) it allows for open date ranges (not covered by ISO 8601).

## 6 Time coverage in the metadata

### 6.1 dc:coverage

xpath: `/c:CMD/c:Components/c:OLAC-DcmiTerms/c:coverage`

Following the best practice in Dublin Core encoding, we use either a *date* or a *date range* here. Note that dc:coverage is overloaded and can also contain spatial information. The VLO importer needs to separate them. A possible separation mechanism is to look whether the content of the dc:coverage element matches (by pattern matching) a date or a date range and treat it as time coverage in these cases.

All other cases are than handed over to the recognition of spatial coverage.

dc:coverage is re-implemented as component cmdi-dc-coverage [http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c\\_1271859438215](http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c_1271859438215)

### 6.2 DC-2502

<http://www.isocat.org/datcat/DC-2502>

Time coverage. This is dc:coverage restricted to temporal coverage. Described and explained as “The time period that the content of a resource is about. Dublin Core definition: The temporal topic of the resource.”

### 6.3 DC-3664

<http://www.isocat.org/datcat/DC-3664>

Time coverage. Described as “Basic time coverage information.”

For practical purposes equivalent to DC-2502?

### 6.4 DC-4343

<http://www.isocat.org/datcat/DC-4343>

Intervall. Described as “a : a space of time between events or states”

For practical purposes equivalent to DC-2502?

This data category is used in the component TimeCoverage [http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c\\_1349361150714](http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c_1349361150714)

### 6.5 DC-5742

<http://www.isocat.org/datcat/DC-5742>

End time. Described as “Time at which an event ends”.

P.S. I was unable to locate a corresponding begin/start category.

## **6.6 DC-3654**

<http://www.isocat.org/datcat/DC-3654>

Start range. Described as “The beginning of a range.”

## **6.7 DC-3655**

<http://www.isocat.org/datcat/DC-3655>

End range. Described as “The end of a range.”

Obviously, DC-3654 and DC-3655 correspond to each other. These two data categories are used in the component TimeCoverage [http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c\\_1349361150714](http://catalog.clarin.eu/ds/ComponentRegistry/?item=clarin.eu:cr1:c_1349361150714)