

CKCC Letter Metadata

Walter Ravenek
July 20, 2012

Introduction

The CKCC corpus consists of a number of correspondences. For the analysis of the letters in the correspondences a set of metadata of good quality is essential. This note describes the metadata currently used, and the way in which it is encoded in TEI.

Correspondences

Each letter belongs to a correspondence which has a unique identifier, as follows:

| <i>Correspondence</i> | <i>identifier</i> |
|---|-------------------|
| Caspar Barlaeus (1584-1648) | barl001 |
| Isaac Beeckman (1588-1637) | beec002 |
| René Descartes (1596-1650) | desc004 |
| Hugo de Groot (1583-1645) | groot001 |
| Constantijn Huygens (1596-1687) | huyg001 |
| Christiaan Huygens (1629-1695) | huyg003 |
| Antoni van Leeuwenhoek (1632-1723) | leeu027 |
| Dirk Rembrandtsz van Nierop (1610-1682) | nier005 |
| Jan Swammerdam (1637-1680) | swam001 |

Letters

The *required* metadata for the letters are as follows:

Identifier

A persistent identifier that is unique within the correspondence the letter belongs to. Currently we use identifiers that reflect the (paper) edition the letters are obtained from.

The concatenation of the correspondence identifier and the letter identifier gives an identifier that is unique within the CKCC corpus.

Date Sent

The date the letter was sent, using the Extended Date/Time Format (EDTF) used by the Library of Congress (www.loc.gov/standards/datetime/). The Gregorian calendar is used.

Name Sender

Person names are referred to by a unique identifier in a reference list of persons. Currently we use a person database that is maintained by Huygens ING.

A letter may have multiple senders.

Name Recipient

Person names are referred to by a unique identifier in a reference list of persons. Currently we use a person database that is maintained by Huygens ING.

A letter may have multiple recipients.

Sender Location

Location names are referred to by a unique identifier in a reference list of locations. Currently we use a location database that is maintained by Huygens ING.

Recipient Location

Location names are referred to by a unique identifier in a reference list of locations. Currently we use a location database that is maintained by Huygens ING.

The *optional* set of metadata is as follows:

Main Language

The language is represented by its IANA code (www.iana.org/assignments/language-subtag-registry). For multilingual letters a list of languages may be given.

TEI Format

In TEI representations of letters we use a new element `meta`. The metadata of a letter is contained in a set of meta elements that is present in the `teiHeader` element, as follows:

```
<TEI>
<teiHeader>
<meta type="id" value="0084"/>
<meta type="date" value="1650-08-02"/>
<meta type="sender" value="huyg003"/>
<meta type="recipient" value="huyg007"/>
<meta type="senderloc" value="denha004"/>
<meta type="recipientloc" value="?"/>
<meta type="language" value="fr"/>
</teiHeader>
<text>
...
</text>
</TEI>
```

Values that are unknown are encoded with a question mark (?).

Notes

The draft EDTF specification currently has revision date June 24, 2011. Of the features we currently support the following:

- 001 (date): `yyyy`, `yyyy-mm`, `yyyy-mm-dd`
- 004 (interval): `yyyy/yyyy`, `yyyy-mm/yyyy-mm`,
`yyyy-mm-dd/yyyy-mm-dd`
- 101 (uncertain/approximate): `yyyy?`, `yyyy~`, `yyyy-mm?`, `yyyy-mm~`,
`yyyy-mm-dd?`, `yyyy-mm-dd~`