DiXiT meets CIDOC-CRM

Reported by Georg Vogeler (Graz University)
Scholarly Digital Editing (SDE)

- **Textual Criticism** since Erasmus of Rotterdam, established research methods in literary and historical studies („editorial sciences“):
  Critical reflection on the historical and medial dimension of „texts“ as core objects of literary studies and sources of historical research
- Making texts „**accessible**“ (establish, publish, comment ...)
- Scholarly discussion on SDE since the 1990s
  - Reference books: Sahle 2013, Pierazzo 2015
CIDOC-CRM in SDE and Projects

Close By

- Blumenbach online (http://blumenbach-online.de/, Wettlaufer et al. *DSH* 2015)
- Representations of Peace (IEG Mainz) (http://www.pax-westphalica.de/, Große / Wagner *DHd* 2017)
- Symogih (http://symogih.org/)
- Ligatus Research Centre (http://www.ligatus.org.uk/lob)

=>

- Modelling „factual content“ (persons, places, events named)
- Modelling physical object
People in the Past and their (Economic) Activities

Archival Documentation (Accounting Documents)

Scholarly Edition

Lists, Tables, Spreadsheets

Humanities Scholars, in particular Historians

RDFs, OWL: bookkeeping.ttl

EAD

Digital Images

TEI

XSL, XQuery

RDF

csv, xlsx, SQL ...

SPSS, R, Excel

Word, PDF, HTML, SVG, ...
CIDOC-CRM and Other Standards

- **FRBR** as important conceptual model for scholarship on texts and of scholarly editing:
  - Editing an expression of a work documented in items producing a new manifestation
  - FRBR(oo) and CIDOC-CRM?
- TEI as most effective standard for production of SDE
  - Conversion to RDF takes CIDOC-CRM as target model (XSL-Stylesheets by Sebastian Rahtz) for persons/places and similar
A TEI document, or a fragment thereof, can be modelled as a CIDOC-CRM conceptual object.

This can be used to explicitly document elements of the CIDOC-CRM model.
TEI as part of CIDOC-CRM

- **E52** "April 16-17, 1742"
- **E31** Schnitler's border protocols Vol. 1, sec. 1.2.
- **P4** has time-span
- **P70** is documented in
- **P128** is carried by
- **P7** took place at
- **P87** is identified by
- **E84** The div element with xml:id="Schn1_6096" of the file schnitler-tei.xml on my disk.
- **E48** "Røros"
Visions for modelling

• Object metadata
  • Text bearing objects => <E84-Information-Carrier> ?
  • „Features“ => creation, change in ownership, modification
  • Reference to „text“ => <P128>

• Textual metadata
  • Texts, versions, works
    • „text“ => <E33-Linguistic-Object> ?
    • “version” => “text as written manifestation” => <E34-Inscription> ?
    • “work” => <E73 Information object> ?
  • Elaborate on „content“: text as an assumption about a real world fact, but the relationship between text and real world is result of an human interpretation (by the editor) („factoids“)
  • Elaborate on other textual properties: textual structure, rhetorics, style, explicit and implicit coreference, ...
Digital Representation of „Editable“ Objects

- Visual perspective
  - crm:E36: visual item
  - crm:E26: physical feature
  - crm:P56: bears feature
  - crm:P138i: has representation

- Linguistic perspective
  - crm:E33: linguistic object
  - crm:P128: carries

- Material perspective
  - crm:E1: Entity
  - crm:P67: refers to

Object (document, text)
Ontologies and perspectives for digital scholarly editions (fellow perspective)

• many paleographers each modeling their domain (writing) in their own way. Sharing a common understanding of the domain using the same concepts and names among researchers would facilitate data interchange and collaboration.

• editing projects with similar or even the same research scopes, all starting from scratch building their data models. Ontologies enable the sharing and reuse of knowledge.

• an edition, marked up in TEI, which represents information as it appears on the document, failing to include references to i.e. persons. Sometimes there may be variant spellings for the same person, or same spellings of a given name which refers to more than one person, sometimes the reference will be implicit, and so forth. This information need to be modeled in a way that the computer understands who is meant and where it refers to.

• if a user of an edition is searching for a place but doesn’t know the exact name of that place. A highly structured and enriched ontology could enable a facetted search approach which limits the number of possible results (‘all cities with more than 350.000 inhabitants in Bavaria’).

• an editor who knows a lot about his material but has difficulties ordering this knowledge and revealing relations of concepts in it. Ontologies can also serve as models of knowledge, enhancing our understanding of the correlations within a certain domain.

CIDOC-CRM and Digital Editions: The Future

- Interchange between TEI as standard for text description and CIDOC-CRM as standard for objects and factoids represented in the text (crm2tei)?
- Clear concept of status of references between a linguistic object and facts („factoid“)?
- Effective software solutions for CIDOC-CRM modelled parts of SDE (put efforts in Erlangen-CRM?)
- Using CIDOC-CRM as upper level ontology for a future ontology of SDE?
Digital Scholarly Editing and CIDOC-CRM

- a perfect couple?

Georg Vogeler
georg.vogeler@uni-graz.at