La ‘Syntaxe du codex’ and TEI. Models, mappings and visualization tools

10.5.2018 Hamburg
A. Beta maṣāḥǝft: Manuscripts of Ethiopia and Eritrea
Collaborations and metadata interoperation with other projects

(Applications)

Ethio-SPaRe manuscripts, images, institutions
Full integration in BM
Encyclopaedia Aethiopica Authority files, persons, places, taxonomy
IHA Islam in the Horn of Africa manuscripts, repositories, literary works
Full integration in BM of all contents
DASI Database of Ancient South Arabian Inscriptions inscriptions
metadata replication
EMIP Ethiopic Manuscript Imaging Project manuscripts, scrolls, images
Metadata and images replication
EMML Ethiopian Manuscript Microfilm Library and Gunda Gunde Manuscripts collection manuscripts metadata
BM reuses metadata and links to images
EMA Ethiopic Manuscript Archives manuscripts, authority files, schema for diplomatic annotations
Full workflow and data integration in BM

Individual collaboration by Jacopo Gnisci schema, guidelines and data encoding of Art Historical elements
Individual collaboration by Eliana Dal Sasso schema, guidelines and data encoding for Binding Descriptions

Collaborations and metadata interoperation with other projects

(Bibliography)

Secondary Bibliography

Show attestations
Select one of the keywords listed from the record to see related data
select...

No keyword selected.

Load hypotheses.Is public annotations pointing here
Use the tag "RehMasino0000078" in your public hypotheses.Is annotations which refer to this entity:

Administrative position
settlement
settlement
region

Quires Distribution for the 7 manuscripts in this selection which have a collation with quire descriptions

8 leaves 2 leaves
7.8% 14.8% 63.3%
6 leaves 5 leaves
10 leaves 1 leaves
4 leaves 3 leaves
2 leaves 1 leaves
Multiple editions

other versions

click on word to open search in dictionary
critical apparatus
contents and physical description for each unit

direct link to each text, placement, clickable text

(you have seen this and the following slide on Monday)
Paris, Bibliothèque nationale de France, BnF Éthiopien 165

Edited by Dunyasha Raiak

Dated

General description

Fragments of chants, Exodus, the Pauline Epistles, Prayers, Weëdëse Mëryâm

Number of Text units: 11

Codicological Unit 1

Origin of codicological unit 1

Before 1877

Contents

Fragment of a chant book for feasts (f.) / Item p1_1
ff. 1-6 Fragment of a chant book for feasts (Dagga = Cae 3178 = C 3178 = work?)
Language of text: Gôzô

Catalogue Bibliography of codicological unit 1


Codicological Unit 2

Origin of codicological unit 2

Before 1877

Contents

Exodus 20:12-21 / Item p2_1
f. 7 Exodus 20:12-21 (Exodus = Cae 1367)
Language of text: Gôzô

Catalogue Bibliography of codicological unit 2

graph view taken directly from TEI

All dates available in the record

explicitly stated relations
B. Beta maṣāḥǝft and *La Syntaxe du Codex*
XML
XQuery
Graphs
RDF
SPARQL
AHEAD!!
Starting needs

- Encode that something happened rather than something can be observed
- This part is later than this other part
- The quire was added inside this quire later
- This part of manuscript comes from another manuscript
- This manuscripts is made of several pieces of other manuscripts
- etc.

“mettre en relation ces éléments pour comprendre comment le codex ‘tient ensemble’ et se modifie avec le temps”
(Andrist, Canart, and Maniaci 2013, 9)
Steps

1. decide on the workflow
2. translate *La Syntaxe* to a formal ontology
3. produce annotations from manuscript descriptions in TEI using the entities and relations in the ontology
4. visualise the annotations as suggested in *La Syntaxe*
5. Examples TEI to Visualization and back again
Principles

• The technology should support
  • speed
  • laziness
  • collaboration
  • complexity
  • I do not know
  • data not available
  • I do not want to say
  • maybe...

• The data models and standard used should support diversified input over time (i.e. quite often and as a default incomplete data, sometimes even erroneous)

• The logic and scripts should make as little assumptions as possible at each stage and if they do make assumptions those should be stupid enough not to generate clever mistakes

• Needs to be reusable, based on widely used and supported standards and using technologies which are as software independent as possible.

• Publishing and using Linked Open Data means we can do thing with other data which we do not yet know. And they can do with our data things we do not yet know. We are making connections we do not even know about.
1. Workflow - La Syntaxe du Codex

STEP 1 = List discontinuities (observation)

- Add to a flat list
- Draw a table where converging discontinuities (discontinuités convergentes) can become visible

STEP 2 = Enrich the table (further research)

with chronological and geographical information in order to verify the relevance of the discontinuities and consequently recognize production units and circulation units (see below).

Make an hypothesis.

STEP 3 = Go back to the Manuscript

check theoretical results and hypothesis with archeological analysis

Data from observation can be encoded in TEI and used to produce the table

OR Data in TEI can be transformed to RDF to produce the table from that format

Data can be encoded in TEI and used to enrich the table

The hypothesis on the identifications of UniProd and UniCirc and their relations are much better represented by a graph, as in the book, so RDF.

To be able to iterate the process in one workflow I need to encode in TEI the relations
1. Workflow - La Syntaxe du Codex

Enter observations and hypothesis in TEI

Convert it all to RDF

produce visualisations

STEP 1: the cataloguer encodes in TEI both description and hypotheses

STEP 2: the cataloguer checks visualizations (table, graphs)

STEP 3: verifying, good if there are images linked in the table

These are produced from the RDF!

- one could produce the same RDF and make the same visualisations using data from a db and not TEI
- the RDF can be reused for many other applications
2. La Syntaxe du Codex as Ontology in OWL

- declare a class for each Unit type and transformation
- declare relations between them
- already all done in the book!
3. Using TEI to encode structural analysis

- there are elements which could be taken one to one to describe a Unit: `<layoutNote>` = UniMep or `<handNote>` = UniMain

- UniMat might already be different, as the material of a part is usually indicated inside `<support>` which is nested in the `<physDesc>`, so could have a different scope

- UniProd and UniCirc can be assigned looking at the TEI programmatically only in very few cases, e.g. the ms being described is always a UniCirc.

- mapping element name to Unit is impossible, as e.g. we use `<item>` both to list additions (UniCont), quires (UniCah) and foliation (UniMarq)

- transformation and their types are nowhere to be encoded
3. Using TEI to encode structural analysis - <locus>

It is essential to be able to produce any tabulation as in the book that there are indications of placement in the manuscript <locus/> is used in BM also to express exact folia and the lines, as well as the corresponding facsimile.
3. TEI to RDF - xml:id structure

To be able to assign URIs dynamically and build a graph, the IDs in the TEI file need to be stable and semantically recognisable.

<table>
<thead>
<tr>
<th>Element</th>
<th>ID pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>msPart</td>
<td>p\d+{,\d+}*</td>
</tr>
<tr>
<td>msItem</td>
<td>p\d+{,\d+}* _\d+{,\d+}*</td>
</tr>
<tr>
<td>binding decoNote</td>
<td>b\d+</td>
</tr>
<tr>
<td>decoNote</td>
<td>d\d+</td>
</tr>
<tr>
<td>addition</td>
<td>a\d+</td>
</tr>
<tr>
<td>extra</td>
<td>e\d+</td>
</tr>
<tr>
<td>quire</td>
<td>q\d+</td>
</tr>
<tr>
<td>foliation unit</td>
<td>fo\d+</td>
</tr>
<tr>
<td>hand</td>
<td>h\d+</td>
</tr>
<tr>
<td>title</td>
<td>t\d+</td>
</tr>
<tr>
<td>name</td>
<td>n\d+</td>
</tr>
<tr>
<td>edition</td>
<td>ed\d+</td>
</tr>
<tr>
<td>transformation</td>
<td>tr\d+</td>
</tr>
<tr>
<td>Uni\w+ (SdC)</td>
<td>Uni\w+\d+ (SdC)</td>
</tr>
</tbody>
</table>
3. TEI to RDF - URIs

For example in *Beta maṣāḥǝft* encoding of Vatican City, Biblioteca Apostolica Vaticana, Aeth. 1 and the Gospel of Luke (CAe 2713)

- The Biblioteca Apostolica Vaticana Ethiopic Manuscript 1 is a manuscript, the entity [http://betamasahaft.eu/BAVet1](http://betamasahaft.eu/BAVet1)
- The Gospel of Luke in the above manuscript is the entity [http://betamasahaft.eu/BAVet1/msitem/ms_i1.4.2](http://betamasahaft.eu/BAVet1/msitem/ms_i1.4.2)
- The Calendaric note on folio 219r is the entity [http://betamasahaft.eu/BAVet1/addition/a3](http://betamasahaft.eu/BAVet1/addition/a3)
- Additionally, each of these is assigned to a class (and it can be added to as many as one wishes)
  - [http://betamasahaft.eu/BAVet1](http://betamasahaft.eu/BAVet1) is an instance in the class [http://lawd.info/ontology/AssembledWork](http://lawd.info/ontology/AssembledWork)
  - [http://betamasahaft.eu/BAVet1/msitem/ms_i1.4.2](http://betamasahaft.eu/BAVet1/msitem/ms_i1.4.2) is an instance in the class [http://betamasahaft.eu/msitem](http://betamasahaft.eu/msitem) and might be an instance in the class UniCont
  - [http://betamasahaft.eu/BAVet1/addition/a3](http://betamasahaft.eu/BAVet1/addition/a3) is an instance in the class [http://betamasahaft.eu/addition](http://betamasahaft.eu/addition) and might be an instance in the class UniCont
3. TEI to RDF - <relation>

<relation> describes any kind of relationship or linkage amongst a specified group of places, events, persons, objects or other items. 

Module: namesdates — Names, Dates, People, and Places

Attributes:
- @name: supplies a name for the kind of relationship of which this is an occurrence of
  - Status: Optional
  - Datatype: teidata.enumerated
- @active: identifies the ‘active’ participants in a non-mutual relationship, or all the participants in a mutual one.
  - Status: Optional
  - Datatype: 1–∞ occurrences of teidata.pointer separated by whitespace
- @mutual: supplies a list of participants amongst all of whom the relationship holds equally.
  - Status: Optional
  - Datatype: 1–∞ occurrences of teidata.pointer separated by whitespace
- @passive: identifies the ‘passive’ participants in a non-mutual relationship.
  - Status: Optional

Example:
<relation
  active="http://betamasahaft.eu/BAVcerulli37#p1"
  name="http://http://Synthax.du.Codex/ontology#constituteUnit"
  passive="http://betamasahaft.eu/BAVcerulli37/UniProd/UniProd1"/>

<relation
  active="BAVcerulli37#p1"
  name="SdC:constituteUnit"
  passive="BAVcerulli37#UniProd1"/>
4. Visualizing the RDF

Some of the informations are extracted from the TEI encoding, some others need to be stated in relation elements.

the visualization is part of the workflow and needs to:
- dynamically update as more content is added
- be available also when some data is lacking
4. Visualizing the RDF - Table of non concomitant discontinuities

<table>
<thead>
<tr>
<th>cahiers</th>
<th>folios</th>
<th>UniMat</th>
<th>UniMarq</th>
<th>UniCah</th>
<th>UniCont</th>
<th>UniMain</th>
<th>UniÉcri</th>
<th>UniRégl</th>
<th>UniMep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (1-8)</td>
<td>/Mat1</td>
<td>/Mq1</td>
<td>/Ch1</td>
<td>/Ct1</td>
<td>/Mn1</td>
<td>/Éc1</td>
<td>/Rg1</td>
<td>/Mp1</td>
<td></td>
</tr>
<tr>
<td>2. (9-16)</td>
<td>10v med.</td>
<td>11r sup.</td>
<td>Ct1/vide</td>
<td>Ct2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. (17-24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. (25-32)</td>
<td>28v inf.</td>
<td>32v inf.</td>
<td>Mat1/</td>
<td>Mq1/</td>
<td>Ch1/</td>
<td>Ct2/</td>
<td>Mn1/</td>
<td>Éc1/</td>
<td>Rg1/</td>
</tr>
<tr>
<td>gard. ant.</td>
<td>/Mat2</td>
<td>vide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gard. post.</td>
<td>Mat2/</td>
<td>vide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- there are no bars like in /Mat1, Mat1/ because the exact column and line can be indicated
- Quires are treated the same way as other units. Actually only 10% of the descriptions in BM have a collation
- no display yet for guards and binding data
- sometimes the grouping of the rows goes wrong
4. Visualizing the RDF - Enriched Table

click the Enrich button and some data is picked up from the TEI and stucked into the table

<table>
<thead>
<tr>
<th>Quires</th>
<th>folios</th>
<th>UniMat</th>
<th>UniMarq</th>
<th>UniCah</th>
<th>UniCont</th>
<th>UniMain</th>
<th>UniEcri</th>
<th>UniRegl</th>
<th>UniMep</th>
<th>decoration</th>
<th>UniProd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>addition</td>
<td>UniMain</td>
<td>UniEcri</td>
<td>UniRegl</td>
<td>UniMep</td>
<td>decoration</td>
</tr>
<tr>
<td>1v - 94r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>e1</td>
<td>h1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94v - 95r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95v - 97v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98r - 98v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99r - 103r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103v - 105v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106r - 106v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107r - 112v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113r - 113v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quires</th>
<th>folios</th>
<th>UniMat</th>
<th>UniMarq</th>
<th>UniCah</th>
<th>UniCont</th>
<th>UniMain</th>
<th>UniEcri</th>
<th>UniRegl</th>
<th>UniMep</th>
<th>decoration</th>
<th>UniProd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>e1</td>
<td>h1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1v - 94r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>msItem ms i1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>additon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
- The French words: X; e; siècle are written in a recent hand.
- id: e1

**Element Names:**
- msItem
- Psalms of David
- Named Entities: Psalms of David | Terms | References:
4. Visualizing the RDF-Transformation models

There is a property which connects UniCirc and UniProd Classes as Domain and Range.
5. Example 1- Oxford, Bodleian Library, Bodleian Aeth. e.

none of the three <msPart> has an internal date and the dating is in general uncertain, but we know that <msPart> 1 and 2 were added later to the manuscript.
5. Example 2 - Vatican City, Biblioteca Apostolica Vaticana, Cerulli 37

A manuscript part has probably been added to the manuscript.
5. Example 3 - - Oxford, Bodleian Library, Bodleian Aeth. f. 11 (R) and Oxford, Bodleian Library, Bodleian Aeth. f. 12 (R)

both are scrolls containing magic prayers which belonged to the same owner and writes at the end of his description of Oxford, Bodleian Library, Bodleian Aeth. f. 11, "Continuation in no. 91" where n.91 is Oxford, Bodleian Library, Bodleian Aeth. f. 12. The research team has discussed this and convened that the two scrolls where once one.
5. Example 4 - Paris, Bibliothèque nationale de France, BnF Éthiopien 45 and Paris, Bibliothèque nationale de France, BnF Éthiopien 165

Paris, Bibliothèque nationale de France, BnF Éthiopien 45 (BNFet165) **contains leaves detached from** Paris, Bibliothèque nationale de France, BnF Éthiopien 45 (BNFet45)
C. Additional benefits and further potential of Linked Open Data
Further potential of the RDF data - Query data about a UniCirc

```
SELECT DISTINCT * 
WHERE 
{bm:BAVcerulli37 a SdC:UniCirc ;
  dcterms:hasPart ?msPart .
?msPart a SdC:UniProd .}
```

```
SELECT DISTINCT * 
WHERE 
{bm:BAVcerulli37 UniCirc/UniCirc1 a SdC:UniCirc ;
  dcterms:hasPart ?msPart .
?msPart a SdC:UniProd .}
```

```
SELECT DISTINCT * 
WHERE 
{bm:BAVcerulli37 UniCirc/UniCirc2 a SdC:UniCirc ;
  dcterms:hasPart ?msPart .
?msPart a SdC:UniProd .}
```

I can ask about any UniCirc using the same query, no subordinate entities
Further potential of the RDF data - Query data about a UniCirc

```sql
SELECT DISTINCT *
WHERE
{?
    AnyUniCirc a SdC:UniCirc ;
    dcterms:hasPart ?AnyUniProd .
?AnyUniProd a SdC:UniProd ;
    SdC:constituteUnit ?UniProdID .
?AnyTransformation SdC:hasTransformationModel ?model
} LIMIT 25
```

I can ask about all UniCirc then filter by, e.g. date, or material, etc. because BAVcerulli37, is actually the same as BAVcerulli37/UniCirc/UniCirc2, we could filter out those which have a skos:exactMatch relation.
Further potential of the RDF data - Query data about a UniCirc

**which UniCirc have the same ModelCah?**

**XQuery the TEI?**

I could get all those which have a collation whose item always have the same number of leaves….

```xml
collection(/location/of/TEI)//collation[count(distinct-values(descendant::item/dim[@unit='leaf'])) = 1]
```

in BM: 56 out of 1287 which have any collation information at all.

or I could, e.g., tabulate all collations making a CSV

```xml
for $collation in collection($config:data-rootMS)//t:collation[t:list]
let $msID := string($collation/ancestor::t:TEI/@xml:id)
let $repoID := string($collation/ancestor::t:TEI//t:repository/@ref)
let $NumberOfQuires := count($collation/t:item)
let $QuireSizes := for $quire in $collation//t:item return $quire/t:dim[@unit='leaf']/text()
let $quiresizesstring := string-join($QuireSizes, ',')
order by $NumberOfQuires
return $repoID || ',' || $msID || ',' || $NumberOfQuires || ', ' || $quiresizesstring
```

but one could get already to the filtered result…
Further potential of the RDF data - Query data about a UniCirc

which UniCirc have the same ModelCah?

```xml
<relation
  active="BAVcerulli37"
  name="SdC:hasUnitModel"
  passive="SdC:Cah1"
/>

SELECT DISTINCT *
WHERE
{
  ?Anything SdC:hasUnitModel SdC:Cah1.
}
```

in BM: 0 out of all, because nobody has yet added this kind of annotation.
Further potential of the RDF data - get summary information about transformations

what can ever happen to a manuscript sitting on a shelf? Tell me about all transformations

SELECT *
WHERE {
  ?transformation a SdC:Transformation;
  SdC:produces ?UnitProduct;
  SdC:hasTransformationModel ?model .
}

As in any query, with any data model, to get further information the query will become more complex...
Further potential of the RDF data - get summary information about transformations

```sql
SELECT DISTINCT ?AnyUnit ?type ?AnyTransformation ?model
WHERE
{?
   ?AnyUnit SdC:undergoesTransformation ?AnyTransformation ;
   a ?type .
   BIND(STR(?type) as ?t)
   FILTER(strStarts(?t, 'http://Syntaxe.du.Codex/ontology#') )
   ?AnyTransformation SdC:hasTransformationModel ?model
}
LIMIT 25
```

One can get comparable results also if the statements are done slightly differently
Further potential of the RDF data - federated queries

```
SELECT *
WHERE
{

  SERVICE <http://wwwb.library.vanderbilt.edu/exist/apps/srophe/api/sparql>
  {
  }
}
UNION
{
}
}
```

Available data can be joined for specific queries
Vielen Dank

Pietro Maria Liuzzo
pietro.liuzzo@uni-hamburg.de

Special Thanks to
Dorothea Reule and Nafisa Valieva
for proposing and discussing the examples
in this presentation
and to all the team of Beta Masaheft
Hiob Ludolf Zentrum für Äthiopistik

Emergency Slides!
This is not finished, it is work in progress.
Can I reuse it?

Yes please!

https://github.com/BetaMasaheft
Can I contribute?

Yes please!

https://github.com/BetaMasaheft

http://betamasaheft.eu/Guidelines/
Why "not" CIDOC-CRM?

True. But:

You can construct your CIDOC on the fly with a CONSTRUCT query

\[
\text{CONSTRUCT} \\
\{ \\
?\text{transformation} \text{ a} \, \text{<http://www.cidoc-crm.org/cidoc-crm/E11\_Modification> ;} \\
\text{<http://www.cidoc-crm.org/cidoc-crm/P31\_has\_modified> } ?\text{AnyUniCirc; } \\
\text{<http://www.cidoc-crm.org/cidoc-crm/P108\_has\_produced> } ?\text{AnyUni } . \\
?\text{AnyUniCirc} \text{ a} \, \text{<http://www.cidoc-crm.org/cidoc-crm/E24\_Physical\_Man-Made\_Thing> } \\
\} \\
\text{WHERE} \\
\{ \\
?\text{AnyUniCirc} \text{ a} \, \text{<http://Syntaxe.du.Codex/ontology\#UniCirc> ;} \\
\text{<http://Syntaxe.du.Codex/ontology\#undergoesTransformation> } ?\text{transformation . } \\
?\text{transformation} \text{ a} \, \text{<http://Syntaxe.du.Codex/ontology\#produces> } ?\text{AnyUni} \\
\} \\
\text{LIMIT 20}
\]

we use cidoc-crm/E11_Modification for each Transformation and so on, providing an internal mapping

With federated query, knowing what mapping you want to apply, results can be obtained