

The Content of Accounts and Registers in their Digital Edition

XML/TEI, Spreadsheets, and Semantic Web Technologies

DRAFT: Published version in: [Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters, hg. v. Jürgen Sarnowsky, Göttingen 2016](#)

Georg Vogeler, Universität Graz <georg.vogeler@uni-graz.at>

Introduction

This article considers the use of semantic web technologies in the context of everyday historians. It deduces from theoretical considerations needs for the actual implementation of a digital edition. It explains some of the basic concepts of the semantic web more extensively than necessary for the digital humanities scholar already familiar with these technologies. I've argued elsewhere why a digital edition can be considered the best method to publish economic records as historical sources.¹ This argument is based on the insight that historical accounting records can only be understood when their documentary properties are fully considered: Their palaeographical and codicological features ("visual" layer) and the linguistics of the text ("linguistic" layer) contribute to their understanding as a formal representation of the facts recorded ("content" layer).² The following first discusses the drawbacks of reducing digital edition of accounts and economic records to the encoding offered by the TEI. I will compare the text oriented approach of the TEI with other digital representations of accounts that are oriented primarily on the economic facts accounted. The second part of the article discusses the opportunities offered by the usage of semantic web technologies (RDF, RDFs/OWL, SKOS and SPARQL) to encode and expose the content layer of digital editions. I have described elsewhere in more detail my own proposal how a customized XML/TEI transcription can be transformed into a XML serialisation of RDF facts,³ and there are other projects interlacing RDF structures into TEI.⁴ This article focus on an introduction into the semantic web technologies as

* All cited URLs were last accessed at on 5.1.2016 and if possible stored in the repository of archive.org via the „Save Page Now“-service (<http://wayback.archive.org/>).

¹ Georg Vogeler: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?, in: Grenzen und Möglichkeiten der Digital Humanities, hg. v. Constanze Baum u. Thomas Stäcker, Wolfenbüttel 2015, DOI: 10.17175/sb001_007, URL: <<http://zfdg.de/warum-werden-mittelalterliche-und-fr%C3%BChneuzeitliche-rechnungsb%C3%BCcher-eigentlich-nicht-digital-ediert>>.

² See Vogeler 2015a##, 2015b##, 2014##, and basic theoretical considerations in particular in Sahle 2013. Similar thoughts are expressed by Thaller 2012##, Pierazzo 2015, Ciula/Eide 2014##. From a much more pragmatic approach the problem is discussed in Jennifer Stertzer: Working with the Financial Records of George Washington. Document vs. Data, in: Digital Studies 4 (2014) <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/247/308#>.

³ Vogeler 2015a##, hier: http://www.zfdg.de/sb001_007#hd3.

⁴ The Sharing Ancient Wisdoms-Project <<http://www.ancientwisdoms.ac.uk/>> converts references between text fragments into RDF structures: Anna Jordanous / Alan Stanley / Charlotte Tupman: Contemporary transformation of ancient documents for recording and retrieving maximum information: when one form of markup is not enough, in: Proceedings of Balisage: The Markup Conference 2012, 2012 <<http://doi.org/10.4242/BalisageVol8.Jordanous01#>>; Stuart Dunn et al.: Sharing Ancient Wisdoms: developing structures for tracking cultural dynamics by linking moral and philosophical anthologies with their

proposed by the W3C and discusses how they can be applied to historical accounts as a common data model, for the creation of controlled vocabularies, in exposing the content layer over the web, and for querying data aggregated from several sources. The final part of the article exemplifies the whole set of methods on data extracted from existing digital editions of late medieval accounts. The presented in this paper is part the MEDEA activities funded by DFG and NEH.⁵

State of the Art: Text Encoding or Spreadsheet?

In practice the palaeographical/linguistic approach and the content approach to accounting documents usually are handled separately.⁶ The researcher interested in the text created flat transcriptions, usually only documenting palaeographical observations. The researcher interested in the numbers created tables, spreadsheets or relational databases. The research on digital edition has focused on the first approach. The TEI as the major standard to encode historical texts provides a rich vocabulary to mark-up this kind of information in the text.⁷ The editors can encode linear and table styled texts; they can link the transcription to an image of the text; non-standard characters can be encoded with the `<g>` element. A particular problem has been addressed by modifications recently: Economic records can be living documents. In these information changes constantly – when liabilities were solved, when tax payers payed etc. There are codes to describe scribal errors, additions and modifications or change in scribal hands, and since 2012 there are sections in the TEI guidelines dedicated to the genesis of a text.⁸

By using XML/TEI digital edition methods thus cover very well the combination of visual representations in images with palaeographical and pure text encoding. An example for this is offered by the edition of the “Amtsbücher” of the Teutonic Order around 1450 by Cordula Franzke and Jürgen Sarnowsky: They offer a full text transcription with palaeographical observations like

source and recipient texts, paper presented at the DH2012 Hamburg, URL: <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/sharing-ancient-wisdoms-developing-structures-for-tracking-cultural-dynamics-by-linking-moral-and-philosophical-anthologies-with-their-source-and-recipient-texts.1.html>; Charlotte Tupmann u. Anna Jordanous: Sharing Ancient Wisdoms across the Semantic Web Using TEI and Ontologies, in: Analysis of Ancient and Medieval Texts and Manuscripts. Digital Approaches, hg. v. Tara Andrews u. Caroline Macé, Turnhout 2015 (Lectio 1), S. 213-228. The RDF of the project is available at <http://www.ancientwisdoms.ac.uk/library/data/>.

⁵ For more details on the project see <http://medea.hypotheses.org>.

⁶ The digital edition of the George Washington Papers is a good example that this is changing recently: Jennifer Stertzer: Working with the Financial Records of George Washington. Document vs. Data, in: Digital Studies 4 (2014) http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/247/308# .

⁷ See for the following TEI Consortium: TEI P5. Guidelines for Electronic Text Encoding and Interchange, originally edited by C.M. Sperberg-McQueen and Lou Burnard, version 2.9.1, 2015 <http://www.tei-c.org/Guidelines/P5/#>, in particular the chapters on elements available in all TEI documents (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html>), default text structure (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>), characters, glyphs, and writing modes (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/WD.html>), and representation of primary sources (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>).

⁸ See Fotis Jannidis, Elena Pierazzo, Malte Rehbein, Lou Burnard, Gregor Middell, Moritz Wissenbach: Manuscripts SIG: Documents and Genetic Criticism TEI Style, [2010/11], URL: <http://www.tei-c.org/SIG/Manuscripts/genetic.html> for the proposal, which was integrated into the official guidelines in 2012 and especially the section on the transcription of primary sources: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.

deletions and editorial interventions. The economic facts recorded are not encoded.⁹ The reason for this might be that encoding economic facts is not covered by straight-forward-TEI. The TEI offers encoding for numbers and measurements. `<num>`¹⁰ allows to encode textual representation of numbers and convert them into computable numerals encoded in an attribute. An encoding like `<num value="43">xliiii</num>` would thus keep the roman number in the text and hand over the value 43 to a computer calculation. `<measure>` and `<measureGrp>` allow encoding text which represents any kind of measurement. `<measure>` “contains a word or phrase referring to some quantity of an object or commodity, usually comprising a number, a unit, and a commodity name”¹¹ and can be grouped together with other measurements, when they “relate to the same object”¹². Thus a text like *unc. It. XXV g. XI* is easily encoded as `<measureGrp><measure>unc.
I</measure> <measure>t. XXV</measure> <measure>g. XI</measure></measureGrp>`. There are already good examples using this kind of encoding: The digital version of the Edward Arber transcription of the Stationer’s Registers created by Giles Bergel, Ian Gadd, James Cummings and the Bodleian Libraries makes heavy use of the `<num>` element (cfr. Listing #1#).¹³ The same is the case with the prototype digital edition of the Fortune Theatre Records.¹⁴

```
<seg type="fee" rend="roman-numerals aligned-right"><!--processing: xijd--><num type="totalPence" value="12"><!--orig: xijd--><num type="pence" value="12">xij<hi rend="superscript">d</hi> </num> </num> </seg>15
```

Listing #1#: Example from the Stations Registers Online showing a possible use of the TEI `<num>` element¹⁶

With this the TEI guidelines offer a generic element for numerals and measurements to serve many purposes. The example in Listing #1# shows that monetary values can be expressed with the `<num>` elements as well as with the `<measure>` element. This flexibility is typical for the TEI and there are many situations in which this feature has to be considered an advantage. But when using it in a specific domain can create more work than necessary: When dealing with accounts money is usually the major commodity measured. The encoding of monetary values in the TEI would be expressed with the help of the attribute `@type`, and the examples used in the guidelines suggest using the keyword “currency”. This is a bit troublesome as you have to create verbose code to encode a very often occurring textual structure. But the situation is more complex. The `@type` value “currency” representing monetary measurements is only suggested by its continuous use in all the examples of

⁹ Amtsbücher des Deutschen Ordens um 1450. Pflegeamt zu Seehesten und Vogtei zu Leipe, bearb. v. Cordula Franzke u. Jürgen Sarnowsky, Göttingen 2015 (Beihefte zum Preußischen Urkundenbuch 3) online at: http://www.v-r.de/de/amtsbuecher_des_deutschen_ordens_um_1450/t-1/1036073/ with references to the TEI files <http://www.v-r->

[wwwspace.de/downloads/dokumente/Amtsuecher/Amtsuecher_DO_um_1450_TEI_OF_186a.xml](http://www.v-r-) and <http://www.v-r->

[wwwspace.de/downloads/dokumente/Amtsuecher/Amtsuecher_DO_um_1450_TEI_OF_186.xml](http://www.v-r-).

¹⁰ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-num.html>

¹¹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-measure.html>

¹² <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-measureGrp.html>

¹³ The Stationer’s Registers Online (Edward Arber transcription) in TEI/XML, ed. by Giles Bergel, Ian Gadd, James Cummings and the Bodleian Libraries, 2013, URL: <http://dhox-dev.nsms.ox.ac.uk/project/stationers-register-online-sro>. See for this and further explanations on the TEI the contribution of James Cummings to the MEDEA Workshop in Regensburg, 24.10.2015: <http://medea.hypotheses.org/314> and the presentation at <https://prezi.com/bjepuzxdgzpl/encoding-financial-records-in-tei-two-case-studies/>

¹⁴ by the Records of Early English Drama (REED), < <http://ereed.cch.kcl.ac.uk/home/>>

¹⁵ https://raw.githubusercontent.com/gbergel/SRO_Arber_data/master/arber1-2014-01-12.xml

¹⁶ https://raw.githubusercontent.com/gbergel/SRO_Arber_data/master/arber1-2014-01-12.xml

the TEI guidelines. In fact the definition of the attribute reads much more generally: “@type specifies the type of measurement in any convenient typology”.¹⁷ The user of the TEI encoded texts thus cannot infer from the TEI guidelines or the formal definition if a measure-element encodes money, spatial extension or the amount of commodities.

To help in this situation the TEI offers the Roma tool and the ODD mechanism.¹⁸ Roma is a web service in which the user can generate a customisation of the TEI reusing existing elements, modifying them and adding new elements in a way that they are consistent with the general definitions of the TEI.¹⁹ The web service writes a formal description of the modifications following the guidelines of the TEI for the documentation of XML schemas, the so called ODD (“One Document Does it all”). The Roma web service can compile a full XML schema (either following the W3C recommendations for XML schemas or in RelaxNG syntax) or a DTD which can be used to help with encoding. This mechanism made for example the encoding of the annual accounts of the city of Basel between 1535 and 1611 effective. The transcribers in the project used abbreviated codes for the amounts booked: <sr:lb>, <sr:sh> and <sr:d> expressed the currency divisions in pound, shilling and pence. In the publication process this encoding was transformed into full TEI encoding as <measure type=“currency” unit=“lb”> etc.²⁰

A major achievement on the path of TEI customisation is the work of Kathryn Tomasek and Syd Baumann. They suggested enhancing the TEI mark up with mark-up dedicated to economic transactions.²¹ Their so called “Transactionography” describes economic facts as a transfer of services, goods and money between actors. It wraps the <measure> encoding in <transaction> and <transfer>. The actors are referenced in attributes @fra and @til. It is basically meant to be a kind of stand-off annotation as it allows referencing the source of the information in @source attribute, but if text consists only of measures and glosses it could be integrated in a transcription.

The proposal by Tomasek and Baumann avoids another issue in encoding historical accounting and economic records with the TEI: They separate the textual from the logical representation, as it is not clear what the basic text structure of bookkeeping entries in the TEI could be. The history of written accounting and the social anthropologic research on literacy has demonstrated that the topological organisation of the information on the page is determined by culture as well by practical need. Modern spreadsheet software suggests that table style accounting is the main visual method for the organisation of economic facts. But the history of accounting shows that there were many visual representations used.²² The TEI offers at least two elements to encode the textual structure of accounts: the table and the list. The two can be considered as special expressions of the other: The table can be considered as a multidimensional list and the list as a table with only one column. If you

¹⁷ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-measure.html>

¹⁸ See <http://www.tei-c.org/Guidelines/Customization/index.xml> for a detailed discussion on the customization methods offered by the TEI.

¹⁹ <http://www.tei-c.org/Roma/>

²⁰ *Jahrrechnungen der Stadt Basel 1535-1610 – digitale Edition*, hg. v. Susanna Burghartz unter Mitarbeit von Sonia Calvi, Lukas Meili, Jonas Sagelsdorff und Georg Vogeler, Basel/Graz 2015 <<http://gams.uni-graz.at/srbas>>, here <http://gams.uni-graz.at/context:srbas?mode=projekt#kodierungsstandards>.

²¹ Kathryn Tomasek and Syd Bauman, « Encoding Financial Records for Historical Research », *Journal of the Text Encoding Initiative* [Online], Issue 6 | December 2013, Online since 22 January 2014, connection on 23 December 2015. URL : <http://jtei.revues.org/895> ; DOI : 10.4000/jtei.895

²² See for example the German territorial accounts studied by Mark Mersowsky ### or the development of tax accounting in the German territorial states in the late middle ages: Vogeler ###.

delve deeper into the concepts representing the structure of the very accounting texts surviving, the focus on these two concepts shows its reduced heuristic value.

1. The TEI table for example is organised similar to the HTML table as a sequence of rows. Each row contains a sequence of cells. However, when using tables for calculating the accountant often enough focused on the column, as they summed up the numbers per column. The logical structure of the table can thus be considered as a two dimensional matrix – and in fact the XML scheme for spreadsheets for the open source spread sheet software gNumeric uses only cells as core a data structure and allocate each cell in the two dimensional grid with the attributes @Row and @Col for row and column.²³

2. High medieval writers considered the sequence of entries to be calculated in a rubric as a block of text consisting of single sentences.

Img. 1#: Rubrikenrechnung#

The TEI would suggest for this a `<p>`aragraph. If you follow bookkeeping practice through time, it is easy to conceive the next step of the evolution again as a `<p>`aragraph: each entry is written in one block (fig. ##).

Img. 2#: Absatzbuchungen#

In this kind of accounting layout the TEI would probably suggest a `<div>` for the logical organisation of the entries. You might even find friends of encoding this as a list – and the definition of the TEI covers this approach as well. Here is not the place to discuss the semantics of the TEI definitions of these elements in detail. All of them are appropriate for the encoding of accounting texts – and all of them exclude essential information from the encoding, i.e. the logical function of the entries: The text is a written representation of economic facts.

This follows the multiple perspectives on text digital editions have to deal with: text can be considered a visual entity (“image”), a trace of the writing act, a linguistic structure (“language”) or a representation of facts (“content”, “data”).²⁴ The TEI as described above offers very useful tools and concepts to encode text as image and text as language. When it comes to the content of the text, the TEI is less far developed. If you encode the accounting facts as in Listing #1#, you are completely in line with the TEI guidelines. The encoding states that somebody paid a fee of 12 pence. This information is conveyed by attribute values: `type="fee", type="currency", unit="pence"`. Thus, the major accounting information, the “data” represented by the text, is not defined by the TEI itself.

In fact many projects publishing data from historical economic records choose to organise the “content” of the documents in databases or spreadsheets. Good examples of this approach are two projects editing documents on trade. The “Donauhandel” project converts the entries of several 17th and 18th century toll registers in Austrian cities along the Danube into a database. The project extracts information about persons and trade goods.²⁵ Additionally the user can browse through the

²³ David Gilbert: The Gnumeric File Format, 2001 <<http://www.jfree.org/jworkbook/download/gnumeric-xml.pdf>>.

²⁴ See note 2.

²⁵ Projektwebseite: <<http://www.univie.ac.at/donauhandel/>>. Zum Projekt: Peter Rauscher, Andrea Serles u. Beate Pamperl: Die Kremser Waag- und Niederlagsbücher. Bedeutung und Möglichkeiten der digitalen Erschließung von wirtschaftshistorischen Massenquellen, in: Pro Civitate Austriae N.F. 17 (2012), S. 57-82,

images of the registers used. It is built by the model of the Sound Toll project which does the very same with the rich documentation from the ships passing the Danish Sound between 1497 and 1857.²⁶ You can search in normalised data on ships and their masters, cargo and taxes paid. Both use relational databases as their back end. They organise their data in tables (e.g. for the Sound Toll project: passage, cargo, taxes and images) in which the original entries are split into fields (e.g. for the passage table the date and the shipmaster's name and domicile, for the cargo table the commodity itself, its ports of departure and destination and the tax paid per commodity) and which are connected by a common identifier. Some of the data in the tables follows an orthography normalized by the editors.²⁷ This kind "structured data" can be encoded in SQL dumps or - less verbose - in comma separated value text files representing a table.

When downloading the data tables from the Sound Toll Online project it becomes clear that it is not easy to reuse this kind of encoding: The tables and fields represent the logic that the researcher extracted from the source. The question arises if these structures could be mapped to a common language. In the business world the eXtensible Business Reporting Language (XBRL) is on the road to establish itself as a data exchange standard, in particular because several national authorities require taxation data in XBRL based format.²⁸ Like the TEI it uses XML to encode the data. It breaks the data down into simple economic facts that are described by pairs of concepts and values in a specific context. The facts are collected in an instance. The concepts are defined by taxonomies. XBRL offers a dedicated data type for monetary data²⁹ and when using the "Global Ledger" taxonomy provided by the XBRL consortium you find basic accounting concepts entry (`gl-cor:entryHeader` and `gl-cor:entryDetail`), amount booked (`gl-cor:amount`, `gl-cor:signOfAmount`), debit and credit (`gl-cor:debitCreditCode`).³⁰ While TEI has established itself as a common encoding semantics for texts in their linguistic and physical representation, XBRL is on the way to become a common reference for the description of economic facts. For the purpose of digital editing accounting texts in their multiplicity as visual, linguistic and economic objects the separation of technologies for textual and factual modelling and encoding remains.

The Semantic Web

This separation of technologies can be considered part of the distinction between conceptual modelling and modelling for production.³¹ Semantic Web technologies have been proposed as a

online auf der Webseite „Der Donauhandel“ unter <http://www.univie.ac.at/donauhandel/wp-content/uploads/Kremser-Waag-und-Niederlagsbuecher.pdf>.

²⁶ The project is online at <http://www.soundtoll.nl/>. Most recently on the project see: Jan Willem Veluwenkamp: Die 'Sound Toll Registers Online' als Instrument für die Erforschung des frühneuzeitlichen Ostseehandels, in: Wiegen - Zählen - Registrieren. Handelsgeschichtliche Massenquellen und die Erforschung mitteleuropäischer Märkte (13.-18. Jahrhundert), hg. v. Peter Rauscher u. Andrea Serles, Innsbruck 2015 (Beiträge zur Geschichte der Städte Mitteleuropas 25), Innsbruck 2015 (Beiträge zur Geschichte der Städte Mitteleuropas 25), S. 365-385.

²⁷ Jan Willem Veluwenkamp: Sound Toll Registers online. Concise source criticism, in: Soundtoll registers, 2011, S. 1f. URL: <http://www.soundtoll.nl/images/files/STROpdf.pdf>.

²⁸ <https://www.xbrl.org/the-standard/>

²⁹ http://www.xbrl.org/Specification/XBRL-2.1/REC-2003-12-31/XBRL-2.1-REC-2003-12-31+corrected-errata-2013-02-20.html#_5.1.1.3.1

³⁰ <http://www.xbrl.org/taxonomy/int/gl/2015-03-25/plt/case-c/gl-cor-content-2015-03-25.xsd>

³¹ Øyvind Eide, « Ontologies, Data Modeling, and TEI », Journal of the Text Encoding Initiative [Online], Issue 8 | December 2014 - December 2015, Online since 01 January 2015, connection on 22 December 2015. URL : <http://jtei.revues.org/1191> ; DOI : 10.4000/jtei.1191.

method to integrate text and data representation. Their application is usually focused on the reference between named entities in the text and real world entities represented digitally in authority files.³² Historical accounts and economic records can demonstrate that the possibilities of semantic web technologies reach further.

The idea of the semantic web introduced by Tim Berners Lee and the W3C is to exchange not only documents over the internet but to share structured data as well. The W3C developed a framework of standards which help to describe data structures and offers standardised query interfaces. The technical specifications of this framework are called “Resource Description Framework” (RDF). It includes a basic formalisation of data as simple graphs consisting of nodes (“resources”) connected by edges (“properties”). On top of this the definitions of RDF allow describing data schemas and formal ontologies with “RDFs” (RDF Schema) or the Web Ontology Language “OWL”. A third set of specifications under the recursive acronym SPARQL (SPARQL Protocol And RDF Query Language) describes how to query RDF data via the internet. It includes a syntax for queries and a specification of functionalities that the so called “SPARQL endpoints” – i.e. interfaces to RDF databases – should offer.³³

Initially RDF was introduced to create structured metadata to web documents. The basic format of RDF is thus a statement about a web resource. These statements are grammatically very simple as they consist of nothing than an URL for the resource as subject, a property of the resource as predicate and a text or another resource as object. This basic model fits very well into the situation of digital edition of accounting records: The “content” of the accounting document can be considered as a statement about the (linguistic) text. Every XML element in a TEI transcription can be identified via the URL fragment identifier syntax³⁴ with a URL. This URL consists of the document URL and a fragment identifier behind the hash-symbol (#): `http://gams.uni-graz.at/o:srbas.1535#bs_Gipszoll-div-0` leads to the section of the 1535 Basel account that carries a `<div>` mark up and the Identifier “bs_gipszoll-div-9” expressed in a `xml:id`-attribute.³⁵ As I have argued elsewhere³⁶ that this basic structure can be combined with vocabularies describing the relationship between text and data like the “p67 Refers to” property of the CIDOC-CRM³⁷ or the “hasTarget” property from the Open Annotation vocabulary.³⁸ Semantic web technologies thus can help to add stand-off annotations to the XML encoding.

³² A good example of this is Arianna Ciula, Paul Spence and José Miguel Vieira: Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project, in: *Literary and Linguistic Computing* 23,3 (2008), p. 311-325 DOI: 10.1093/lc/fqn018. See also recently an example extending the reference to artifacts and concepts: Jörg Wettlaufer, Christopher Johnson, Martin Scholz, Mark Fichtner and Sree Ganesh Thotempudi: Semantic Blumenbach. Exploration of Text–Object Relationships with Semantic Web Technology in the History of Science, in: *DSH* 30,1 (2015), p. 187-198, DOI: 10.1093/lc/fqv047.

³³ ##Autoren und Titel# <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> and related documents like <http://www.w3.org/TR/rdf-schema/> for RDFs, <http://www.w3.org/TR/rdf-sparql-query/> for SPARQL and <http://www.w3.org/TR/owl-ref/> for OWL. An overview over the activities of the RDF working group of the W3C can be found at http://www.w3.org/2011/rdf-wg/wiki/Main_Page

³⁴ ### , p. 16 <<https://tools.ietf.org/html/rfc3986#section-3>>.

³⁵ A more generalized approach is the XPointer Framework (<http://www.w3.org/TR/xptr-framework/>) which allows the use of XPath expressions in the URL.

³⁶ ### http://www.zfdg.de/sb001_007#hd3

³⁷ Definition of the CIDOC Conceptual Reference Model, produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group, ed. by Patrick Le Boeuf, Martin Doerr,

This is very much in line with the further development of RDF, as this simple but highly structured encoding helps to exchange structured data over the web without being too restrictive on the form of the data. This is in particular encouraged by W3C definition of schema descriptions in the RDF framework. They allow defining the predicates and organising data into classes. Single data points could therefore be assigned to general definitions – or as it would have been called in tables: column headers can be defined explicitly and the schemes can be exchanged via the web. A RDF download of accounting data could thus (a) reference explicitly a transcription of the text and (b) add a formal description of the data structure used to extract economic information from the transcription. Already this recommends RDF for the digital representation of the “content” layer of historical accounting documents.

There are several methods to code the RDF and to handle RDF data. In my personal work practice I’ve made good experiences with the Turtle notation³⁹ and blazegraph⁴⁰ as database. In the context of TEI encoded transcriptions it seems reasonable to prefer RDF/XML as serialization as it can be created with pure XSLT application to TEI transcriptions.⁴¹ The practical decisions in this field do not affect the general argument, as fortunately all serialisations express the same basic concepts and there is software available to convert from one to the other and that the W3C specifications guarantee that RDF data could be stored and queried in all RDF enabled database solutions.

In fact the RDF solution carries advantages beyond the practical decision for one or the other encoding of the data representing the “content” of the accounts and economic records. I’ve already mentioned the first: It is easy to model with RDF the relationship between the textual resource and the data representation. The second is that you can write a formal description of a data model you want to apply. With RDFs you can organise your data structure into a class hierarchy and you can define a set of properties linking between the resources and defining literal descriptions to resources. The economic data represented in the text of accounting records could thus be described in a formal way which allows software engineers to build efficient search engines and statistical data analysis. A third advantage is that RDF is conceptualized as a directed graph i.e. a relationship between two entities which has a start and an end. The transfer of money or goods as a basic concept of accounting can easily be understood as directed graph: one person (start entity) transfers money/goods (relationship) to the other (end entity). Semantic web technologies offer therefore currently the best methods to describe not only the reference between text and economic facts but to describe a generic structure of these facts.

A major interest of scholarly editing is to make available reliable texts to other research. The same is valid for digital editing. In the case of economic records this means that editions should the “content” layer should also be available for research not foreseen in the original setting. While the

Christian Emil Ore, Stephen Stead, Version 6.2 May 2015, S. 67, Online: <http://www.cidoc-crm.org/cidoc-crm/#67>.

³⁸ W3C Open Annotation Data Model, ed. by Robert Sanderson, Paolo Ciccarese, Herbert Van de Sompe, Community Draft, 08 February 2013, here: <http://www.openannotation.org/spec/core/core.html#BodyTarget>

³⁹ <http://www.w3.org/TR/turtle/> ##. Recently the JSON-LD notation (<http://www.w3.org/TR/json-ld/>) is gaining popularity due to the fact that it enables the JSON data representation that well established among programmers to be used as linked data and RDF serialization.

⁴⁰ <http://blazegraph.com/>.

⁴¹ See http://www.zfdg.de/sb001_007#hd3 and <http://gams.uni-graz.at/context:srbas?mode=projekt#kodierungsstandards> for a discussion of an example of this conversion and

digital representation of palaeographical and linguistic knowledge about the document can refer to the TEI as a description standard the data has not found a common point of reference yet. A basic RDF scheme for the description of data to be found in this kind of documents is needed.⁴²

An RDF model can start with the definitions of transactionography. When reading the definitions of transactionography it becomes clear that XML in this case is used to serialise this kind of abstract ontology not to mark-up text. In fact transactionography has already developed a simple ontology for accounting facts: a *transaction* between two *parties* or *accounts* consists of at least one *transfer from* one *to* the other. It transfers a *measurable* and can be *attested* by text. The transfer occurs at a *place*. Booking a transfer into an account can create *liabilities held by* a party and *owed to* another.⁴³ From the XBRL taxonomy one can add the specialization of measurements as *monetary values*. XBRL-GL offers additional data types applicable to the encoding of historical accounting documents: The *entry* is an information fragment of a transfer often only naming one party as the other can be deduced from the textual context of the entry. When it writing down into the ledger, *debit* and *credit* are coded e.g. by using the appropriate columns in a table. Accounts can be compared in a *balance*. XBRL in general adds the concept of calculated values like *totals*. This is in particular useful when encoding historical accounting document. They usually convey these calculated values (e.g. per page, per rubric, per accounting period) which could be compared to values calculated by the computer and thus show transcription or encoding errors viz. calculation errors of the clerks.

Following the TEI model measurements has at least three properties: *what* is measured, what kind of *unit* is used, and in which *quantity* it is measured. In RDF the type of measurement would better be a description of the unit used not of the individual measurement.

Economic history often is not only interested in the accounting practice and the original accounting structure but in economic facts reported in the accounts. Prices, wages and currency conversion rates are good examples for this research interest. The model should therefore contain entities to describe the relationship between two measurable as *prices* and should identify measurable not only as monetary item but as *commodities* or *services* as well. Prices can be considered as a special case of *conversions* of measurements.

All of this can be expressed in RDFs and here is not the place to explain this formalisation further. An advantage of using RDF for the modelling can be shown with the concept *account*. It is a problematic term in the vocabulary because it can have different meanings: In accounting theory an account can be “personal”, “real”, or “nominal”. It therefore can represent a relationship between two partners, an aggregation of items of value, or a temporary evaluation of the economic status of business. But accounts can also be considered as a list of bookkeeping entries.⁴⁴ In accounting on paper they could be expressed in single pages or dedicated ledgers and they are identified by a heading. In final fair copies of bookkeeping documents in particular in the context of single entry bookkeeping they are just lists of entries under a heading which could be called “rubrics”, referencing the medieval method of marking headings with red ink. Fortunately there is a common abstraction in the two concepts when you consider “account” as method to organize a group of single facts. In the XBRL scheme this

⁴² This could fulfill the dreams of Michel Morat, who asked for a unique “program” to aggregate the information from different port accounts, Michel Mollat: *Comptes portuaires et informatique*, in: *Informatique et histoire medievale*, Rom 1977, p. 149-156, here p. 154 and 156.

⁴³ <http://www.customization.encodinghfrs.org/transactionography.html>

⁴⁴ ##Merriam-Webster##.

is considered a *dimension* of the data.⁴⁵ Modelling this in RDFs allows organising the information in a hierarchy of abstractions: Both *account* and *rubric* can be described as special case of *dimension*. The *account* is an abstraction of *personal account*, *real account* and *nominal account*.

Expressing this model in RDF should enable projects with different research interests to link their data model a common abstraction. Let me consider this with the example of `##Sound toll##` and `##`:

The Sound Toll records STR report the passage date, the name of the shipmaster, his place of residence, his port of departure and – from the mid-1660s – his port of destination, the composition of the cargo and the toll paid. Each of these entries is thus a *bk:Entry* made referring to a specific *bk:date*. The original purpose of the registers was to record the *bk:MonetaryAmount* of taxes received. The entry conveys information on commodities (*bk:Commodity*) in the cargo and can be analysed under different dimensions (*bk:Dimension*): the shipmasters residence, the ports of departure and destination of the cargo. In double entry bookkeeping you could consider the shipmaster as a *bk:PersonalAccount*, which would be a specialised interpretation of the general concept of *bk:Dimension*.

In this general system you can identify single transactions: Identified by the number 830801 the clerks of the Sound Toll (*transfer to*) receive on July 22nd in the year 1600 (*date*) *from* Franndtz Moller from Rostock the *amount* of 16.5 daler and 13 skilling. This tax is payed for the total value of 1688 daller of his cargo.

```
@prefix bk:<http://gams.uni-graz.at/rem/bookkeeping/#> .
```

```
@prefix str:<http://soundtoll.nl/#> .
```

```
@prefix tei:<http://www.tei-c.org/ns/1.0/#> .
```

```
str:830801 a bk:Transfer ;
  bk:transfers [
    a bk:MonetaryValue ;
    bk:quantity 16.5 ;
    bk:unit str:daler ],
  [
    bk:quantity 13 ;
    bk:unit str:skilling ] ;
  bk:from str:M_F_R_1600204 ;
  bk:to str:RoyalDanishCustoms;
  bk:date "1600-07-22" ;
  bk:attestedBy <http://www2.soundtoll.nl/sonttolregisters-042/Sonttolregisters-042_0219.jpg> ;
  bk:entryDetail str:830801-1 .
str:M_F_R_1600204 a bk:personalAccount ;
  tei:persName "Franndtz Moller" ;
  str:from str:Rostock .
str:830801-Price-0 a str:Tax;
  bk:priceOf
    [ a bk:MonetaryValue ; bk:quantity 16.5 ; bk:unit str:daler ] ,
    [ a bk:MonetaryValue ; bk:quantity 13 ; bk:unit str:skilling] ;
  bk:priceFor [ a bk:MonetaryValue ; bk:quantity 1688 ; bk:unit str:daler
  ] .
```

⁴⁵ ###

Listing #2a#: Fragment of Sound Toll Register data expressed in the bookkeeping vocabulary proposed at <http://gams.uni-graz.at/rem/bookkeeping>

The entry states further economic information: The tax is paid for a specific cargo. This consists of six different commodities and the entry gives to each of them a value. Listing #2#b gives an example of one of these entry details.

```
str:830801 bk:entryDetail str:830801-1 .
str:830801-1 a bk:Entry ;
  bk:amounts [
    bk:quantity 16 ;
    bk:unit str:lester ;
    bk:what str:Rugh ] ;
  bk:hasDimension str:Rugh .
str:830801-Price-1 a bk:Price ;
  bk:priceOf [
    a bk:MonetaryValue ;
    bk:quantity 704 ;
    bk:unit str:daler ] ;
  bk:priceFor [
    bk:quantity 16 ;
    bk:unit str:lester ;
    bk:what str:Rugh ] .
str:Rugh a bk:Commodity .
```

Listing #2b#: Fragment of Sound Toll Register data expressed in the bookkeeping vocabulary proposed at <http://gams.uni-graz.at/rem/bookkeeping>

What is obviously missing is the relationship between the ship owner, the passage and the commodities. As the basic economic ontology does not provide a vocabulary for this, it seems appropriate to create a local ontology: `str:Passage str:by <FrantztMoller-from-Rostockh>` ; `str:transports [a bk:Measurable]` ; `str:departure <Danzig>` ; `str:through <DanishSound>` – and in the period before 1634 the *destination* of the goods is usually not mentioned in the Sound Toll registers. It would certainly make sense if other projects editing trade documents agree on a common vocabulary to describe their data. For a research on prices the common `bk:Price` class associated with the goods transported gives already an easy access to the data and the evaluation at the Danish Sound could be easily compared to the market prices in Rostock and Danzig.

The textual evidence for the data would be encoded by `bk:attested`. This leads to an archival identification of the document from which the information is extracted. The link to the image or to a full transcription of the text would probably better be a generic `cidoc-crm:p67` -relationship to the archival document as well.

The 75 annual accounts of the city of Basel from 1535 to 1610 do not use highly sophisticated accounting methods. Each of them organises the income and expense of the city in about 160 rubrics and draws a final total. In the digital edition of these accounts created in cooperation between a team at Basel University under the lead of Susanna Burghartz⁴⁶ and the Centre for Information Modelling at Graz University this records were transcribed and organised in single `bk:Entry`. Each of them reports one `bk:MonetaryAmount`. The entries can be dated (`bk:Date`) roughly by their

⁴⁶ Jahrrechnungen der Stadt Basel 1535-1610 – digitale Edition, hg. v. Susanna Burghartz unter Mitarbeit von Sonia Calvi, Lukas Meili, Jonas Sagelsdorff und Georg Vogeler, Basel/Graz 2015 <<http://gams.uni-graz.at/srbas>>.

attribution to the fiscal year and they are organized into rubrics (*bk:Rubric*). In fact, the rubrics have a structure which is not completely expressed in the written text. There is a group of rubrics concerning the city to be distinguished from the rubrics recording income and expenses from the hinterland. But only the entries from the hinterland have a common heading: *So ist inn uszern schloszen und empteren empfangen*. To encode this data structure the digital edition adds a “virtual rubric” `bs:StadtEinnahmen` for the city. Additionally some entries were assigned to “virtual rubrics” following business evaluation criteria: The accounts report cash flow and thus do not distinguish between money expended as loans viz. received from debtors as repayment of these loans on the one hand and money received by taxes and trade viz. expended on services and goods on the other. This accounting method creates outliers in a numerical evaluation (fig. #1#) which can be avoided by the virtual classification.

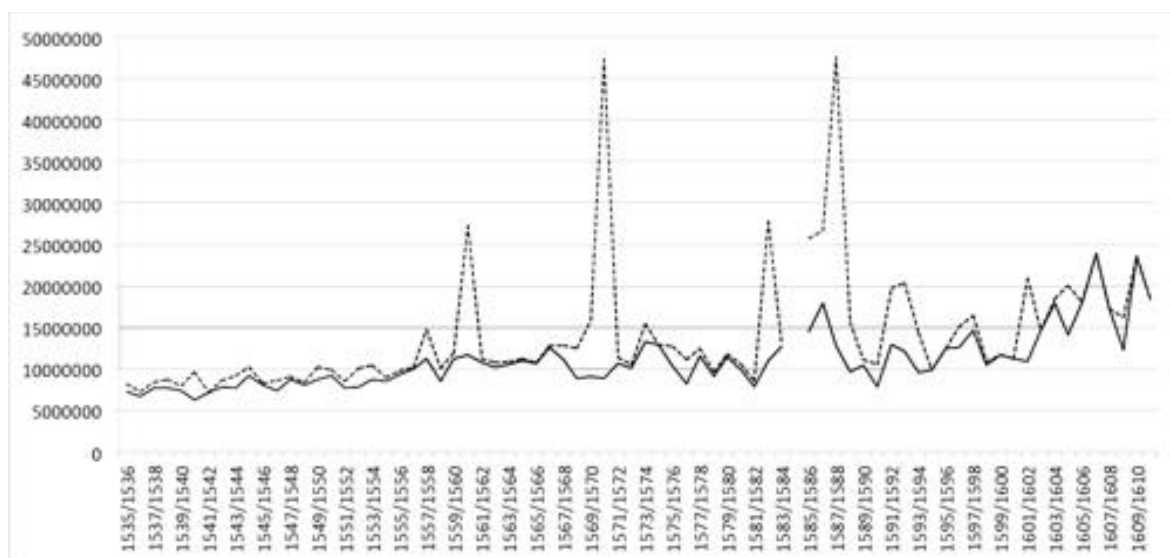


Fig. #1#: Geldzufluss und Gesamtertrag (in d), in: Jonas Sagelsdorff: *Jahrrechnungen der Stadt Basel Erläuterungen, Struktur, Aufbau*, Oktober 2015 (<http://gams.uni-graz.at/context:srbas?mode=rechnungen>), Abb. 2

Both examples show that both types documentation share common features of data structures although they contain very different information. Representing these common features in RDF still allows organising the data for the different research interests.

But what does it help? Indeed the basic organisation of economic facts reported in historical accounting documents makes the economic data interchangeable only at a very basic level. A comparison between the value of wine in Basel and in the Baltics can only be achieved if entries documenting wine can be identified in the Swiss annual accounts and in the Danish toll registers. The data model of the Sound Toll online already helps to achieve that as the commodities in the cargo of the ships passing the Sound are stored in the database. The project has published a multilingual list of these commodities to facilitate search.⁴⁷ You can find there normalised terms and translations which gives the reader at least 57 terms identifying wine: Red wine, white wine, sweet wine, Rhine wine, dry wine from Malaga, wine from Croatia, Cap wine etc. The digital edition of the accounts (“debt- and invoice-registers”) of the Teutonic Order around 1400, created by Cordelia Heß, Christina Link and Jürgen Sarnowsky offers an extensive index. It includes a hierarchical classification of goods,

⁴⁷ ## <http://www.soundtoll.nl/index.php/en/onderzoek/productenlijst>

offices, professions etc.⁴⁸ The Schuredo taxonomy for groceries includes a list of nine types of wine: *Elsasser wyn* (Alsatian wine), *Gobenysch wyn*, *Garschonier wyn/wyn von Gorsgonien* (Gascogne wine), *lantwyne* (rural wine), *Osey wyn*, *reynval* (wine from Riviglio), *rynisscher wyen* (Rhine wine), *Romanya wyn*, *welschis wyn* (Italian wine). With Semantic web technologies the Rhine wine in both lists could be compared, the information on it merged and queried as a whole, as they allow to create a common vocabulary for both sources, the exposure of the information as RDF data and the aggregated query with SPARQL.

The common vocabulary could be expressed in the system the W3C has developed to describe lists of concepts as taxonomies, the "Simple Knowledge Organisation System" (SKOS). Representing the taxonomy from the Prussian edition in SKOS is for example a straight forward task.⁴⁹ In SKOS each of the terms in the taxonomy are considered an abstract concept (*skos:Concept*) that can have several labels to name them (*skos:prefLabel*, *skos:altLabel*). The terms are identified by URIs and are part of a concept scheme. SKOS can express hierarchical relationships between concepts as "broader" and "narrower" or just undefined relations.

With this data description standard the digital edition of accounting records can enhance the "content" layer. In entity-attribute modelled data bases this is usually done by normalisation of text and by offering lists the researcher can use to prepare calculations and queries. Semantic web technologies convert this into formal controlled vocabularies. Controlled vocabularies abstract from the name in the text allowing multilingual research, and they describe hierarchical relationships. With semantic web technologies therefore common index lists for commodities and services, places and persons, or currencies and measurements can be created.

Although it is technically possible it remains an open question from the conceptual point of view, if a vocabulary for commodities and services covering a wide chronological and geographical realm can be achieved at all. But even if not, SKOS can establish identifications even between separated lists: The Rhinewine listed in the Sound toll registers as "Vin, Rhinsk" and as "Rhinskvin" can be identified to be an exact match to the Rhinewine in the registers from the Teutonic order as "rynisscher wyen" identified in the classification as "Rheinwein" under the URI <http://www.schuredo.uni-hamburg.de/browse/groceries/0023/0023.A/0007>" (listing #3#).

```
<skos:Concept rdf:about="http://soundtoll.nl/products#Rhinskvin">
  <skos:inScheme rdf:resource="http://soundtoll.nl/products"/>
  <skos:prefLabel xml:lang="dk">Rhinskvin</skos:prefLabel>
  <skos:altLabel xml:lang="dk">Vin, Rhinsk</skos:altLabel>
  <skos:prefLabel xml:lang="en">Rhine wine</skos:prefLabel>
  <skos:prefLabel xml:lang="nl">Rijnwijn</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Vin du Rhin</skos:prefLabel>
  <skos:prefLabel xml:lang="nl/frs">Rynwyn</skos:prefLabel>
  <skos:exactMatch rdf:resource="http://www.schuredo.uni-
hamburg.de/browse/groceries/0023/0023.A/0007"/>
</skos:Concept>
```

Listing #3#: Mögliche SKOS-Repräsentation des Produktregisters der
Sundzolldatenbank online

⁴⁸ <http://www.schuredo.uni-hamburg.de/content/main/classifications.xml>

⁴⁹ See a sample file at ###.

As Tomasek/Bauman have pointed out that controlled vocabularies created for modern taxation purposes like the Harmonized System established by the World Customs Organization⁵⁰ fail when it comes to historical commodities.⁵¹ Problems of socioeconomic classification of historical records have been discussed long.⁵² It seems therefore appropriate to create RDF representations of commodities and services from existing editions of historical texts like. The list of the Sound toll registers or the SchuReDo classification are good examples for the possible starting points. For occupations the Historical International Standard Classification of Occupations has done extensive work. The classification is based on the modern scheme ISCO68 but adds micro classes extracted directly from historical records. The work group is extending it for socioeconomic status and other related domains.⁵³ These resources can be enhanced with general data resources like dbpedia⁵⁴, wiki data⁵⁵ or the lexical databases following the “Wordnet”-model.⁵⁶ For monetary data numismatics researchers have started building an ontology of numismatic concepts.⁵⁷ Currently it is focused on ancient denominations but offers a basic ontology to build on for medieval and early modern coins.

Another particular useful resource for the description of the “content” layer of historical accounting records are common measurement conversions. Projects like the “Global price and income history working group”⁵⁸ link to a small selection of global conversions which are partially available as spreadsheets. Usually this kind of information is currently extracted from printed source and resources.⁵⁹ The concept of digital edition adding a “content” layer can integrate these texts into the framework of historical economic facts represented in the semantic web. In fact, the basic ontology of economic facts described above can be applied the digital edition of reference works for historical metrics. Listing #4# gives an example for a this TEI encoding for a fragment of Nelkenbrechers Taschenbuch from the 14th edition from 1828⁶⁰ in which the text describes the capacity measurements used in Frankfurt/Main. The manual had to explain the partition of the old and a new capacity measurement used in particular for wine and the conversion of both of them into other measurements as for example the French metrical system.

⁵⁰ <http://www.wcoomd.org/en/topics/nomenclature/overview.aspx>

⁵¹ ###Tomasek / Bauman 2013##, p. 3.

⁵² See for example Helga Schultz: Probleme sozialökonomischer Klassifikation, in: Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung, hg. v. Manfred Thaller, St.Katharinen 1986 (Historisch-Sozialwissenschaftliche Forschungen 20), S. 179-185.

⁵³ Marco H.D. van Leeuwen, Ineke Maas and Andrew Miles, *HISCO: Historical International Standard Classification of Occupations*. Leuven 2002. The database is online at <http://historyofwork.iisg.nl/>. The work is continued as a collaborative project: <https://collab.iisg.nl/web/hisco/about>.

⁵⁴ <http://dbpedia.org>.

⁵⁵ <http://wikidata.org>

⁵⁶ The English: <http://wordnet.princeton.edu/> ### RDF representation: <http://wordnet-rdf.princeton.edu/>. See for the concept and other examples (like the GermaNet: <http://www.sfs.uni-tuebingen.de/GermaNet/>, or in the multilingual alignments in the Open Multilingual Wordnet: <http://compling.hss.ntu.edu.sg/omw/>): Lothar Lemnitzer u. Claudia Kunze: Computerlexikographie, Tübingen 2007, S. 135-142.

⁵⁷ <http://nomisma.org/>

⁵⁸ <http://gpih.ucdavis.edu/>

⁵⁹ There is as vast amount of literature from the 18th and 19th century supporting acting trade, e.g. the at least 20 editions of Johann Christian Nelkenbrecher’s Taschenbuch der Münz-, Maß- und Gewichtskunde für Kaufleute printed between 1763 and 1890.

⁶⁰ Johann Christian Nelkenbrecher: Allgemeines Taschenbuch der Münz-, Maaß- und Gewichtskunde für Banquiers und Kaufleute, überarbeitet durch Johann H. Bock u. Carl Crüger, Berlin ¹⁴1828, S. 130f.; online unter <http://reader.digitale-sammlungen.de/de/fs1/object/display/bsb10292452_00150.html?zoom=0.7500000000000002>

“Das Getränkemaß ist zweierlei, nämlich Alt- und Neu-, auch Jungmaaß. Der Unterschied zwischen beiden erstreckt sich bis zur ganzen Maaß hinauf; als 1 Ohm hat 20 Viertel, 80 alte und 90 neue Maaß. Die alte und neue Maaß sind die Einheiten des Getränkemaßes, und jede der beiden Maaße wird in 4 Schoppen eingetheilt. 1 Fuder Wein ist 6 Ohm, und 1 Stück Wein 8 Ohm. (...) Es sind also 100 Frankf. Aichmaaß = 179,272 Franz. Litres (...) und 100 Frankf. Jungmaaß = 159,352 Franz. Litres (...).“

```
<p>Das Getränkemaß ist zweierlei, nämlich <rs ref="ffm:Maaß-alt">Alt-</rs>
und <rs ref="ffm:Maaß-neu">Neu-</rs>, auch <rs ref="ffm:Maaß-
neu">Jungmaaß</rs>. Der Unterschied zwischen beiden erstreckt sich bis zur
ganzen Maaß hinauf; als <seg ana="rem:Conversion"><measure
type="rem:Capacity" unit="ffm:Ohm" quantity="1">1 Ohm</measure> hat
<measure type="rem:Capacity" unit="ffm:Viertel" quantity="20">20
Viertel</measure>, <measure type="rem:Capacity" unit="ffm:Maaß-alt"
quantity="80">80 alte</measure> und <measure type="rem:Capacity"
unit="ffm:Maaß-neu" quantity="90">90 neue Maaß</measure></seg>. Die alte
und neue Maaß sind die Einheiten des Getränkemaßes, und jede der beiden
<seg ana="rem:conversion"><measure type="rem:Capacity" unit="ffm:Maaß"
quantity="1">Maaße</measure> wird in <measure type="rem:Capacity"
quantity="4" unit="ffm:Schoppen">4 Schoppen</measure></seg> eingetheilt.
<seg ana="rem:Conversion"><measure type="rem:Capacity" unit="ffm:Fuder"
commodity="rem:Wine" quantity="1">1 Fuder Wein</measure> ist <measure
type="rem:Capacity" quantity="6" unit="ffm:Ohm" commodity="rem:Wine">6
Ohm</measure></seg>, und <seg ana="rem:Conversion"><measure
type="rem:Capacity" unit="ffm:Stück" quantity="1" commodity="rem:Wine">1
Stück Wein</measure> <measure type="rem:Capacity" unit="ffm:Ohm"
quantity="8">8 Ohm</measure></seg>. <gap/> Es sind also <seg
ana="rem:Conversion"><measure type="rem:Capacity" unit="ffm:Maaß-alt"
quantity="100">100 <rs ref="ffm:Maaß-alt">Frankf. Aichmaaß</rs></measure> =
<measure type="rem:Capacity" unit="L" quantity="179.272">179,272 Franz.
Litres</measure></seg> <gap/> und <seg ana="rem:Conversion"><measure
type="rem:Capacity" unit="ffm:Maaß-neu" quantity="100">100 <rs
ref="ffm:Maaß-neu">Frankf. Jungmaaß</rs></measure> = <measure
type="rem:Capacity" unit="L" quantity="159.352">159,352 Franz.
Litres</measure></seg> <gap/>.</p>
```

Listing #4#: Extract from Nelkenbrechers Allgemeines Taschenbuch ¹⁴1828, p. 130f.:
Transcription and TEI encoding with references to RDF definitions from the economic facts ontology (prefix: *rem*) and a hypothetical specific ontology of measurements in Frankfurt/Main (prefix: *ffm*).

An RDF representation of the Sound Toll Registers as well as of the Basel account would thus best be complemented by conversion tables for the monetary units used, taxonomies for the commodities in

the cargo of the ships, gazetteer identifiers for the places named⁶¹ and internal identifiers for the persons in their different spelling from entry to entry⁶².

This example makes clear that the potential of the formal description of basic data structures and the creation of controlled vocabularies in RDF unfolds in particular when the data is exposed online and when the data can be reused in different environments. The Semantic Web project of the W3C groups the according methods under the term “Linked Open Data”.⁶³ A commonly used controlled vocabulary is one part of the linkage between open data resources. But the example above shows that the reused information can be extracted from digital editions. In fact accounting documents are a major source for information on conversions in particular on prices and wages. Instead of publishing lists separated from the source the usage of semantic web technologies in digital editions integrates the information with the text. The “content” layer of the digital edition becomes the data resource needed for economic analysis.

Technically semantic web technologies offer two methods to aggregate data from different resouces: You can copy public RDF data you’re interested in into your SPARQL engine with the SPARQL instruction “LOAD”⁶⁴ or you can address an external SPARQL engine via its “SPARQL endpoint” in a federated query with the SPARQL instruction “SERVICE”.⁶⁵ From practical point of view both methods have drawbacks: The first method has to be executed before the actual query is done. Additionally, the SPARQL instruction “LOAD” is recent and thus not every SPAQRQL engine supports it, although support is increasing. Similar solutions are offered by OpenLink’s middleware “URIBurner” that is shipped together with the Virtuoso triple store⁶⁶ or the mechanisms to resolve FROM clauses in the Apache Jena ARQ engine⁶⁷. The second method can be inserted in a single query, but has to rely on the performance and the reliability of the provider of the SPARQL endpoint.⁶⁸ For the editor, i.e. the data provider, it is sufficient to publish a RDF version of the “content” layer of the digital edition in one of the major RDF formats (i.e. RDF/XML⁶⁹, RDFa⁷⁰,

⁶¹ The publicly available data from [geonames.org](http://www.geonames.org) is currently the most widely used resource for this purpose. It does not contain names of places not inhabited today anymore, but projects like the Pelagios project (<http://pelagios-project.blogspot.co.uk/>) start to build the necessary resource. An RDF representation of dictionaries with historical place names would add further useful information. The Orbis Latinus by ##Graesse-Benedicte-Plechl## for example is already part of the iMGH web service:

<http://www.mgh.de/dmgh/imgh/geo/faq>.

⁶² The Sound Toll Registers online provide a very usefull tool to reduce probable variations of the spelling in the registers before 1634 when the clerks entered each shipmaster’s passage in separate books distinguishing by the type of cargo## taxed: # <http://dietrich.soundtoll.nl/public/sv.php?proc=1##>.

⁶³ ##Tim Berners-Lee (2006-07-27). "Linked Data—Design Issues". W3C.

<http://www.w3.org/DesignIssues/LinkedData.html>

⁶⁴ <http://www.w3.org/TR/2013/REC-sparql11-update-20130321/#load>

⁶⁵ <http://www.w3.org/TR/2013/REC-sparql11-federated-query-20130321>

⁶⁶ See <http://linkeddata.uriburner.com/> for more detail.

⁶⁷ <https://jena.apache.org/documentation/query/index.html>

⁶⁸ See <http://sparql.es.ai.wu.ac.at/> for an evaluation of 545 SPARQL endpoints.

⁶⁹ The digital edition of the Basel annual accounts by Susanna Burghartz and her team (2015, <http://gams.uni-graz.at/srbas>) uses this format: <http://gams.uni-graz.at/context:srbas?mode=projekt#die-digitale-edition>.

⁷⁰ See as an example for this approach the editions of the Max Weber Foundation: <http://quellen-perspectivia.net/de/portal/start>.

Turtle/N3 or JSON-LD), and not to setup a SPARQL endpoint as elaborate querying will most probably be executed in an environment controlled by the user.⁷¹

Example

Let me explain the complete setting of a semantic web representation in a digital edition with an example. It extracts information on the economic activities of millers in the first half of 15th century from two digital editions of accounts: account of the charter house in Aggsbac and the registers of the Teutonic Order. The account of the charter house in Aggsbach in Lower Austria for the period from 1422 to 1432 has been transcribed by Thomas Aigner. This transcription was converted it into a TEI file,⁷² from which economic facts can be extracted. The ontology on economic facts presented earlier is good enough to cover the basic facts in this account: Entries and measurable transferred can be entered into the text with the support of regular expression identifying the Arabic numbers followed by currency denominators “lb”, “ß”, “d”. The coding of the transactions posted into the register needs manual intervention. The editor has to identify the parties between the money and the goods are transferred. In the entries is named usually only one partner of the transaction, as the other is the charter house itself, identified in the general heading. Converting the TEI into full transaction RDF needs therefore mechanisms to insert the complement partner in each entry. The account names millers at least four times: *molitor* and *molendinum* occur on fol. 30r, 32v and twice on fol. 33v. They are named in a list of liabilities of the monastery, a list of expenses for wine in the year 1434, a sale of a mill in Aggsbach and the expenses by the official of the monastery in Toppel. It is not clear in every case if the term *molitor* describes a profession or is a family name. Thus it makes sense to check for contemporary economic activities of millers. Would they credit money? Could they own a vineyard?

Fortunately millers occur in the contemporary financial registers of the Teutonic Order as well. It is certainly not the first source for answers to the questions above, but it has a great advantage: The digital edition has a very detailed index. In the text they are identified in German as *molmeister* (OF 155, S. 260; OF 153b, S. 73), as *molerynne* (OF 155, S. 292), as *Molner* (OF 155, S. 414; OF 153b, S. 107), and as *wyntmolner/wyntmo^elner* (OF 154, S. 91 u. 92; OF 155, S. 414 u. 415, OF 153b, S. 107). The classification of occupations in the digital edition of the registers gives a clue to all of these occurrences via the abstract identifier “<http://www.schuredo.uni-hamburg.de/browse/profession/0032/0032>”.⁷³ The content of these entries can be encoded with the ontology described above: The millers and wind millers are recorded for liabilities they have with the Teutonic Order in Prussia. They owe money for mill stones and beer and eventually pay their debts. This information is attested by entries in the registers, for some of them even exist multiple evidence in several registers.

Both texts can be linked with the controlled vocabulary of the HISCO. The appropriate code for the work executed by millers is 771.⁷⁴ Introducing it to the Aggsbach account and matching it with the

⁷¹ The SPARQL endpoint provided by the manufacturer of Virtuoso (OpenLink Software) can be used for testing federated queries with RDF files imported from the web as well as querying SPARQL endpoints:

<http://uriburner.com/sparql>.

⁷² ###, preliminary edition by Thomas Aigner, ## <http://gams.uni-graz.at/o:rem.###>.

⁷³ <http://www.schuredo.uni-hamburg.de/browse/profession/0032/0032>

⁷⁴ http://historyofwork.iisg.nl/list_rubri.php?keywords=771

SchuReDo classification leads to a semantic web representation of all the three resource like listing #5#.

```
##Code##
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:bk="http://gams.uni-graz.at/rem/bookkeeping/#"
  xmlns:schuredo="http://www.schuredo.uni-hamburg.de/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept
rdf:about="http://historyofworking.iisg.nl/list_micro.php?keywords=771">
    <skos:prefLabel>Miller</skos:prefLabel>
    <skos:exactMatch rdf:resource="http://www.schuredo.uni-
hamburg.de/browse/profession/0032/0032"/>
  </skos:Concept>
  <bk:Transaction rdf:about="http://www.schuredo.uni-hamburg.de/OF155-
574">
    <bk:attestedBy>Item her Waler, der molmeister czum Elbinghe,
tenetur 6 m. vor 1 l. Wismersch bir[e]. Item 13 m. vor 26 to. b[i]r[s].
Summa 19 m.</bk:attestedBy>
    <bk:consists-of>
      <bk:Transfer rdf:about="http://www.schuredo.uni-
hamburg.de/OF155-574-d">
        <bk:to rdf:resource="http://www.schuredo.uni-
hamburg.de/WalterElbing"/>
        <bk:from rdf:resource="http://www.schuredo.uni-
hamburg.de/Großschäffer"/>
        <bk:transfers>
          <bk:Commodity>
            <bk:quantity
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</bk:quantity>
            <bk:unit>1.</bk:unit>
            <bk:what>Wismersch bir</bk:what>
          </bk:Commodity>
        </bk:transfers>
      </bk:Transfer>
    </bk:consists-of>
    <bk:consists-of>
      <bk:Liability rdf:about="http://www.schuredo.uni-
hamburg.de/OF155-574-c">
        <bk:held rdf:resource="http://www.schuredo.uni-
hamburg.de/WalterElbing"/>
        <bk:owedto rdf:resource="http://www.schuredo.uni-
hamburg.de/Großschäffer"/>
        <bk:transfers>
          <bk:MonetaryValie>
            <bk:quantity
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">6</bk:quantity>
            <bk:unit rdf:resource="http://www.schuredo.uni-
hamburg.de/Mark"/>
          </bk:MonetaryValie>
        </bk:transfers>
      </bk:Liability>
    </bk:consists-of>
  </bk:Transaction>
```

```

    <schuredo:Person rdf:about="http://www.schuredo.uni-
hamburg.de/WalterElbing">
      <rdf:type rdf:resource="http://www.schuredo.uni-
hamburg.de/Person"/>
      <schuredo:occupation rdf:resource="http://www.schuredo.uni-
hamburg.de/browse/profession/0032/0032"/>
    </schuredo:Person>
  </rdf:RDF>

```

Listing #5#: RDF fragments of the accounts from the charter house in Aggsbach, the register from the Teutonic order and the SKOS matching instruction.

It would probably be stored on three different servers: <http://gams.uni-graz.at/o:rem.aggs1422/RDF> contains the RDF representation of the accounts of the charter house of Aggsbach.

<http://www.schuredo.uni-hamburg/OF153b/RDF>, <http://www.schuredo.uni-hamburg/OF154/RDF> and <http://www.schuredo.uni-hamburg/OF155/RDF> would be good URLs for RDF representations of the accounts from the Teutonic Order. <http://www.schuredo.uni-hamburg.de/RDF/profession> could be the link to the occupation taxonomy used in the edition. Its RDF should include skos:exactMatch statements linking the taxonomy to the HISCO taxonomy, which should provide a RDF representation as well.

A query to get an insight into the economic activities of millers in the beginning 15th century could read like listing #6#a. This SPARQL makes use of the above mentioned mechanisms from JENA ARQ or the Virtuoso URIBurner to import external RDF resources into the SPARQL engine. A SPARQL using federated search in SPARQL endpoints to the above mentioned resource would look like listing #6#b.⁷⁵

```

# the economic activities of millers in the beginning 15th
PREFIX hisco:<http://historyofworking.iisg.nl/list_micro.php?keywords=>
PREFIX bk:<http://gams.uni-graz.at/rem/bookkeeping/#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

```

```

SELECT ?entry ?prop ?val
FROM http://gams.uni-graz.at/o:rem.aggs1422/RDF
FROM http://www.schuredo.uni-hamburg/OF154/RDF
FROM http://www.schuredo.uni-hamburg/OF155/RDF
FROM http://www.schuredo.uni-hamburg.de/RDF/profession
WHERE {
  ?entry ?involves ?p ;
    ?prop ?val ;
    a bk:Transfer .
  ?p ?occup ?o .
  #The reference to hisco can be an skos:exactMatch or explicit
  {
    {hisco:771 skos:exactMatch ?o .}
  }
  UNION
  {?p ?occup hisco:771 . }
}

```

⁷⁵ The data from the example has been stored in the GAMS repository of Graz University as demonstrator, see <http://gams.uni-graz.at/archive/objects/o:rem.3890/datastreams/ONTOLOGY/content> for Schuredo data, <http://gams.uni-graz.at/o:rem.aggsb1422/RDF> for the Aggsbach data and <http://gams.uni-graz.at/archive/objects/query:rem.mueller/methods/sdef:Query/getXML> for the SPARQL result

}

Listing #6#a: SPARQL query to search for all entries referring to someone with the occupation “miller” in a RDF representation of the accounts of the charter house in Aggsbach and the registers of the Teutonic Order – all resources in one repository

```
# the economic activities of millers in the beginning 15th
PREFIX hisco:<http://historyofworking.iisg.nl/list_micro.php?keywords=>
PREFIX bk:<http://gams.uni-graz.at/rem/bookkeeping/#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

SELECT ?entry ?prop ?val
WHERE {
  SERVICE http://www.schuredo.uni-hamburg.de/SPARQL {
    ?entry ?involves ?p ;
      ?prop ?val ;
      a bk:Transfer .
    ?p ?occup ?o .
    hisco:771 skos:exactMatch ?o .
  }
  UNION
  SERVICE FROM http://gams.uni-graz.at/rem/SPARQL {
    ?entry ?involves ?p ;
      ?prop ?val ;
      a bk:Transfer .
    ?p ?occup hisco:771 .
  }
}
```

Listing #6#b: SPARQL query to search for all entries referring to someone with the occupation “miller” in a RDF representation of the accounts of the charter house in Aggsbach and the registers of the Teutonic Order – resources access via SPARQL endpoints

Conclusion

Semantic Web technologies were developed to facilitate exchange of structured data over the web. They are the method of choice when it comes to expose the “content” layer of a digital edition of historical accounts and economic records online. Unfortunately only single projects have made use of the technologies. This is probably due to the fact that computer engineers consider SQL databases as the major method to represent and query structured data. An analysis of RDF and SPARQL as the main Semantic Web technologies has made clear that they are able to fulfill# the needs of many historians interested in the “content” layer: RDFs/OWL allows modeling and encoding of the basic economic facts recorded in historical accounts and economic records. SKOS allows describing taxonomies of commodities, services and monetary values recorded. They can be aligned into a common vocabulary. SPARQL allows aggregate querying of resources on these common facts together with individual data recorded.

The next step for projects like the Sound Toll Online, the SchuReDo digital edition or any other digital representation of historical records of economic data is therefore to follow examples like the digital edition of the Basel annual accounts in the 16th century: This edition has started as “simple” transcription of the texts adding information to TEI that allows an easy transformation into basic RDF resources. They represent not the complete richness of the text, as they model only the basic

structure of accounting entry and organization of the entries in rubrics. But as the rubrics add some concepts drawn from modern accounting theory they already explain outliers in the income/expenses balance#: ##Credits repaid##. Digital editions of registers and accounts that try to express their interpretation of the text in a “content” layer and that publish this interpretation online with the help of semantic web technologies thus do what scholarly edition is meant to do: publishing the critical analysis of the document by a competent scholar.