THE FUTURE OF INFORMATION SCIENCES

INFUTURE2019

KNOWLEDGE IN THE DIGITAL AGE

Edited by

Petra Bago, Ivana Hebrang Grgić, Tomislav Ivanjko,

Vedran Juričić, Željka Mikložević and Helena Stublić

Zagreb, November 2019

7th International Conference
The Future of Information Sciences
INFuture2019: Knowledge in the Digital Age
Zagreb, 21-22 November 2019

Organizer
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
University of Zagreb
Croatia

Publisher
Faculty of Humanities and Social Sciences, University of Zagreb
Department of Information and Communication Sciences
FF press

All papers were reviewed by at least two reviewers. INFuture relies on the double-blind peer review process in which the identity of both reviewers and authors as well as their institutions are respectfully concealed from both parties.

ISSN 2706-3518

DOI: https://doi.org/10.17234/INFUTURE.2019

CONTENTS

https://doi.org/10.17234/INFUTURE.2019.9

Original scientific paper

Event-based Modelling of a Major Historical Government Source
Ministerratsprotokolle 1848–1918

Stephan Kurz
Austrian Academy of Sciences, Vienna, Austria
stephan.kurz@oeaw.ac.at

Wladimir Fischer-Nebmaier
Austrian Academy of Sciences, Vienna, Austria
wladimir.fischer@oeaw.ac.at

Summary
Our paper showcases a critical-historical document edition with a long tradition and of considerable size, the "Ministerratsprotokolle" (MRP). We are currently transferring the MRP to a digital-edition paradigm, based on the XML markup scheme proposed by the Text Encoding Initiative (TEI). Our paper starts out by presenting the corpus and discussing the workflows that lead to the present state of the MRP data. Our main task is to edit, but also to disseminate this important digital Cultural Heritage resource. In order to open access to a broader public quickly, our choice fell on the easiest to process and most general category in our code: events. Events in our case include first of all the dates of ministerial council sessions and the agenda items discussed during these sessions. For these two kinds of events, we propose a markup strategy that is compatible to RDF statements, linking documented text and facts which the text is referring to. To insure reusability from across all disciplines, we are using a prototype eXist-db application that serves the data both as TEI XML and via API. The aim of our paper is twofold: To theoretically discuss event-based modelling of textual resources, and to describe the corpus unlocked by this type of modelling.

Key words: digital edition, data modelling, event-based modelling, Linked Open Data, mass data, event, Text Encoding Initiative

Introduction

The *Ministerratsprotokolle der Habsburgermonarchie und Österreich-Ungarns* (MRP) are a major corpus of governmental documents stemming from the Habsburg Monarchy's administrative legacy.[1] Covering nearly sixty years of government, the minutes (protocols) of the Ministerial Council are one of the few edited resources that on the one hand display the inner workings of the Monarchy's governments, and represent a huge data mine full of prosopographic, political, administrative, economic, cultural, and social information in general, on the other. Structurally, the MRP are organised as a series of session events that each include agenda item events.

Our research institution is responsible for the MRP's scholarly edition under the proposition that cultural heritage must be steadily curated to keep it in circulation. To all historians, preserving historical heritage and making it accessible to the public in a scientifically prepared form is what basic research means. For historians in a digital paradigm, the on-line availability and functionality of historical textual resources has become key, as "[h]istorians consider the content of the text 'data', and they want to use this data in their research to gain knowledge about the past." (Vogeler, 2019: 309)

In our paper, we are presenting the data and underlying event-based data model that has been implicit in the MRP edition's text. We want to make this underlying model explicit in order to make the data more accessible to research outside of the historical disciplines. We are discussing two converging, but distinct data sets within our overall edition project: one comes from a large amount of retro-digitized material, the other one derives from our current "hybrid edition" research and editing

[1] Throughout the paper, we will reference the material we are discussing by the siglum "MRP." For general information on the edition series and on the editing guidelines that have not substantially changed since their inception (Rumpler, 1970). For details on the digital edition workflow (Kurz et al., 2019).

process. But before focussing on the editorial workflow and specifically on entities and events implied, allow for a brief overview on the edition's contents and contexts.

The edition corpus and data

The MRP minutes document the Council of Ministers of the Austrian Empire (which included Hungary at the time) from its advent in 1848 up to 1867 (Series 1). The 1867 Austro-Hungarian Compromise transformed the Austrian Empire into the "dual" Austro-Hungarian Monarchy. Three bodies emerged from the hitherto unified Council of Ministers: the Joint Council of Ministers of the Austro-Hungarian Monarchy (Series 2), a Council of Ministers for the Kingdom of Hungary, and one government for the remaining countries, abbreviated as Austria or Cisleithania, all of them spanning the years 1867 to 1918. The minutes of the latter government's sessions are being edited as "The Minutes of the Cisleithanian Council of Ministers, 1867−1918" (Series 3).

The Council of Ministers or *Ministerrat* was the central body of government. The minutes of its sessions reflect all aspects of political life, from issues concerning the state's structure and organization to social, economic and technical developments as well as cultural and social problems. The protocols were journalized by dedicated recording clerks and later on circulated among the participants, before being presented to the emperor for his formal decision ("Allerhöchste Entschließung"), which put their content into immediate effect.

The structure of the minutes is relatively uniform. Each one starts with a list of the members of government present or absent, and of domain specialists invited at certain occasions. Then follows a table of contents and, as the main part, a detailed summary of the propositions and discussions having taken place during the meetings. When particularly controversial topics were on the agenda, the exchange of opinions is well graspable, but usually, the texts are concise summaries of propositions, arguments and outcome. It is important to note that the minutes are not recorded in direct speech, but rather written in the third person.[2]

Who is the document edition good for? Classical historians have thus far used it for intellectual information extraction, mostly with regard to political decision making. But the minutes are full of information both of encyclopedic value and with a potential for quantitative processing. Names of persons involved in political affairs as well as lower employees of the state, extending also to foreign officials, are abundantly present. Attached to those persons are information on their lives and careers, titles and decorations, as well as their relations to other persons. A plethora of institutions are being mentioned, which over time changed both their names and responsibilities. The same is true for place names and regional entities from all over the Monarchy, from present-day Montenegro to Poland, from Ukraine to Italy. Finally, most of the agenda items refer to at least one law or decree from the Austrian legal codes. The language of the minutes and the arguments made are large-scale specimens of late 19th-century administrative German and of the political discourse of Austria-Hungary's elites. And this is only the text of the minutes themselves. The scientific comment in footnotes and the extensive introduction of each volume, enriches all this with references to contemporary news articles, legal codes and to all relevant other minutes connected to the respective agenda item. To sum up, the MRP combine "text from the source with interpretation by and for historians" (Vogeler 2019: 312) and are thus good for classical historians and also for cultural, social, economic, and quantitative historians, as well as corpus linguists, historical linguists and discourse analysts (among others).

All these raw materials that have been buried between book covers (for general notes on related challenges see Piotrowski, 2012), are waiting to be transformed into the largest historical data mine of the political, cultural, legal, economic, and technological history of the Habsburg Empire.

[2] This modal difference keeps us from directly using teiParla, the recently developing standard for parliamentary minutes: Erjavec/Pančur recently organized a workshop for a proposed teiParla standard, see https://www.clarin.eu/event/2019/parlaformat-workshop. At the Austrian Academy of Sciences, T. Wissik will be applying this to the ParlAT corpus of contemporary Austrian parliament debate transcripts.

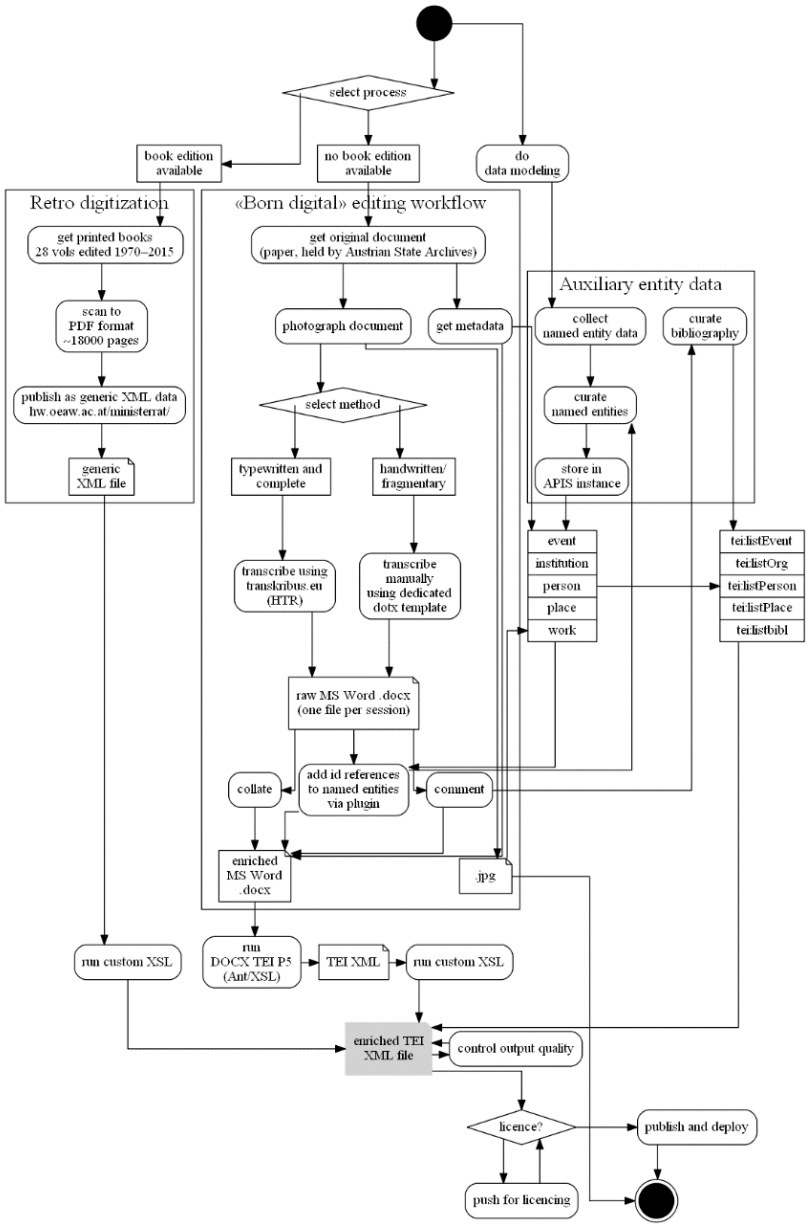

Figure 1. Ministerratsprotokolle 1848–1918: Digitisation Workflow Activity Diagram. Source: own work. Graphviz source files for diagrams are available in the [GH02] repository

In the following, we present the detailed editing process in order to explain the important issues of this edition (Fig. 1, for the deployment infrastructure cf. Fig. 2 below). It starts with "sources," i.e. thousands of more than one-hundred-year-old documents in the Austrian State Archives: handwritten and typewritten minute forms. The minutes of the 1848–1867 period have been transcribed in the past forty years by hand, using a word processor. Editors have then annotated, commented and supplemented them with lists of terms to be used in indices in the appendix. After a manual typesetting step done in Adobe InDesign, editors then highlighted terms intended for the index by felt marker and added them to the index manually, before the books were printed. The resulting 18,000 pages are the data we are retro-digitizing.

From Print to Hybrid

The transformation of the print edition to a full-text source has met some difficulties. The first volumes of the edition were still typeset in the pre-digital age. Subsequently, machine-readable original files of the transcripts were used, but many of them have been destroyed. The same is true for the InDesign files of the more recent volumes. Therefore, the 28 already edited print volumes had to be scanned and re-keyed and have been made available in a generic XML format by the Austrian

Academy of Sciences' in-house publisher. This did not include any additional semantic markup, although the actual content is both structured and provides semantic clues. In other words, the print edition is currently "only" available as flat text data.

To remedy this situation and provide better access to the existing data, our research institution have decided to convert the data into the widely-accepted XML format proposed by the Text Encoding Initiative (TEI Consortium, 2019), for which different tools are already available in order to enrich the data semi-automatically. Further reasons to opt for TEI encoded XML as target format include: It is -) standardized (Bunard, 2019), -) both human and computer readable, -) highly accepted as an exchange and archiving format, -) able to accommodate a superset of our editorial needs, given the limited markup depth that we can afford to produce within the given time constraints, and -) it fits both of our data sets, which saves development effort and enables us to think about the MRP as one common data source. Due to unsolved licencing questions at the Austrian Academy of Sciences Press, we currently cannot publish the TEI full-text derivates of the retro-digitized material. Generic XML data is available at [HW] under legally unclear "Open Access" conditions.

Since 2018, we have been using an XML-based hybrid edition workflow that moves the semi-automatic semantic markup a step closer to the transcription phase. For the three concurrent volumes that we are editing at the time of writing, we are using an MS Word .docx workflow (cf. Fig. 1, 'Born digital editing workflow'). This includes linking entities via a JavaScript plugin and automates part of the structural parameters of the edition text by applying paragraph and character styles context-sensitively.[3] The resulting .docx XML data is then pre-processed through standard OxGarage OOXML .docx to TEI templates. The resulting documents are further customized using XSL transformations, especially regarding genre-specific structural markup and other details. The TEI output of both the "born digital material" and the "retro digitization" pipelines are equivalent, which is important for the next steps of processing and for the deployment of the output.

When assessing the distinct qualities of our edition as a whole, we concluded that we are dealing with a trove of dated (or dateable) facts that have left their traces in the records. Although in the TEI world, the use of the event element has not been very prominent,[4] for our edition, events turned out to be a central structural element, which we use both for the structural organisation of textual elements and for providing new discovery tools on the application level.

What is the background of this? For enriching both the retro-digitized and the born-digital MRP data, we look for automatically detectable items that yield date strings identifiable by regular expression string matching. Those either refer to 1) acts of reference or citation, or to 2) events outside the textual sphere ("facts"). In the first case, depending on the context before and after a date string, we are able to add links to external data sources, such as the digitized newspaper (ANNO) and legal documents (ALEX) archives held by the Austrian National Library, and also to reference points within the MRP corpus itself (such as a mention of preceding or following sessions). In the second case, and even more prominently, we model the minutes themselves as describing events on the level of agenda items: The session event of a particular day contains references to all points on the agenda, which we also understand as events. Each of the agenda items has one or more actors whose presence and utterances are recorded in relation to the agenda item event. Both the events and relations have already been present in the MRP paper edition and digitized full text in the "short regesta" or abstracts, but only implicitly. Taken together, we are in a position to not only hyperlink existing references and dates but to also make the structure of events explicit in the markup of the edition data. Thus, it is safe to say that the whole MRP edition hinges on events.

Entities and Events

Most of the entities can be modelled with obvious and well-used elements of the TEI vocabulary: dates in time, named entities, such as individual persons, geographical places, bibliographical entities such as laws and newspaper articles.[5] Yet, where possible, the mere occurrence of a date has to be properly rewritten as an event to gain meaning.

[3] For examples of work-in-progress .docx documents, cf. http://mrptestapp.acdh-dev.oeaw.ac.at/.

[4] Since 2010, there have been multiple interventions on the TEI mailing list that aim toward referencing events from within the text, but so far, no single solution has been adopted by the TEI community.

[5] MRP related challenges concerning named entities and building indices from them are discussed in Kurz/Zaytseva 2019.

In order to collect the event structure and link the session's agenda item texts to the events via a stable URI for the born-digital part of our data, we are using a relational database and web front-end (dubbed "Auxiliary entity data" in Fig. 1). This is done with an instance of the APIS database system,[6] in which we are creating entries for the event, person, institution, place and work entity types, including authority file identifiers where applicable. For the majority of textual mentions such as person names, this is done semi-automatically, as names of ministers and other government officials are recurring in the texts; other entities do not even exist in any of the usual Linked Open Data sources (WikiData, VIAF/GND, GeoNames); we have to manually add those to our database while editing the source texts.[7]

However, it is a different case with the pre-existing texts from the printed books: We cannot re-edit the textual data from scratch since our time resources are limited. Therefore, we can only apply automatic information extraction strategies one step at a time, and we have opted to start out with events:

Since dated facts are equally common in both our data sets, we attempt to at least match the majority of possible dateable facets, and wrap them in date/@when-iso elements for further analysis. The respective xsl:analyze-string regex solution targets contemporary writing styles for dates, including some abbreviated forms employed in the MRP text that would otherwise need domain specific expertise and/or manual reading (e.g. "1. l. M." refers to the first of the current month ['laufenden Monats'], in this case 1864-07-01).

Although we originally only have a string representation, we can infer that we are indeed dealing with a dated event that is relevant to the MRP corpus, and for which we understand "event" to be a change of conditions related to a subject and an object that took place at a specific point in time. Consequently, any event can be expressed as a subject–predicate–object triple with a date attribute. This is purposefully compatible with a triple-based logic at least for the description of events that are linked to the edition text.[8] Currently, we are only applying this logic to the textual facets mentioned, as they present extra-textual *facts* that form interpretational additions to the underlying *text* on the editor's behalf. Hence, we separate them from the latter: The fact that a council session took place (a particular topic was discussed) is extraneous to the minute's text – it is a (well-sourced) observation by the editor.

In modelling the data structure for the born digital workflow, our starting point were the textual units we refer to as "agenda items." These are the basic units of the digital edition's layout and data model. A ministerial council session is not tied to a certain date in a one-to-one relation, it could be intermitted and last for several days. Therefore, we decided to use the agenda item as the basic unit as it can in almost all cases be tied to a particular singular date. After this decision we could define which TEI element best fits the agenda item. Our choice was to encode the textual content as div type="agenda_item" in the document's body, and to additionally replicate the label assigned to the agenda item in the "Protokollbuch" (book of protocols, a second source that is physically distinct to the actual minutes) into the event element in the profileDesc/abstract/ listEvent of each XML file's teiHeader.

In the model created while transforming generic retro-digitized XML data into TEI, we replicated the session event and the contained agenda item events directly via XPath selection. These are the only events where this is possible without manual intervention with all necessary data regarding the *who*, *what*, *when* and *where*. For *who*, we construct a listPerson with role attributed person elements. *When*

[6] For more on the project during which this Austrian Prosopographical Information System was developed, cf. https://apis.acdh.oeaw.ac.at/. Python and a PostgreSQL database drive its backend, with a Django frontend in place for manual curation and an API e.g. for autocomplete plugins like the one we are using: https://github.com/dariok/officeEntityPlugin.

[7] Named entity recognition and the variety of challenges that come with it is beyond the scope of this paper. Among other strategies, we have been experimenting with the Pelagios Recogito system that uses Stanford NLP tools for additional NER markup, cf. https://recogito.pelagios.org/. For the retro-digitised part of our corpus, named entities that form the base of person and place indices will have to be added at a later stage.

[8] It would even be possible to remodel the edition completely based on RDF triples, as the Swiss "Nationale Infrastruktur f r Editionen – Infrastructure nationale pour les éditions (NIE – INE)" (https://fee.unibas.ch/de/nie-ine/) are successfully proposing. Apart from the necessary resources for such a transition, the humanities environment community, in this case the MRP editors, are favoring a more human-readable TEI approach over a pure LOD approach for the time being. For discussion of the encoding of RDF relationships within the TEI, cf. https://github.com/TEIC/TEI/issues/1860.

and *where* can be inferred by string parsing; they are constructed from the source XML during XSLT processing together with the list of agenda items. The all-important *what* is populated from the session's formal heading in the minutes and the agenda item's description in the "Protokollbuch," respectively.

The Guidelines of the TEI Consortium propose the use of the event element as "data relating to any kind of significant event associated with a person, place, or organization." Thus, it can be placed in the descriptions of said entities, e.g. to list events in a biography that are related to one natural or legal person, or that took place at a particular place. In addition, the element may be used on its own, as long as it is either nested, or grouped in a listEvent container within most of the analytical descriptors the TEI guidelines have on offer. For the MRP edition, we chose to accommodate a listEvent within the teiHeader, wrapped in a profileDesc/abstract construct. This provides convenient stand-off markup that models the event as categorically different from the text that is describing the event; in other words, our approach avoids mixing up interpretation and textual or documentary evidence (Vogeler 2019: 318).

Moreover, our proposed usage of event does not call for any adaptations of the current TEI definition of event.[9] For the time being, we only cover the macro structure of historical facts considered events. There are numerous examples for other events that may be extracted in the future, e.g. we could also lend event status to a text passage giving evidence that a particular minister said something within the scope of one agenda item (of type "utterance," which might include one of type "quotation" etc. ad libitum) like "Der Eisenbahnminister erinnert daran, dass …."

We see the following advantages of our decisions, so far:

- Our model adopts a TEI based workflow for both retro-digitised and newly edited content and thus contributes to the broader aim of making Cultural Heritage available in the digital age through standardisation and reusability.
- Putting events in the centre of our data model allows for easier dealing with the problem of interlinking textual documents and extra-textual facts.
- By using events as hinges between fact and text, we contribute to easier and sustainable accessibility of the documents we are editing, and to the development of new discovery tools.

While implementing the workflow outlined above, we also had to create a working environment not only for data input and curation, but also for the eventual publishing of the output. In Figure 2, we show the deployment components involved.

[9] This converges with the fact that source editions from the historical disciplines tend to use abstracts (short regesta) to sum up the content of the given text. As this practice is not universal across the disciplines, there are efforts to make listEvent more interchangeable while keeping the distinction between evidence and interpretation. A joint paper of one of the authors with scholars of other disciplines has been presented at TEI 2019, cf. the "Recreating history through events" paper by C. Fritze, H. Klug, S. Kurz and C. Steindl, https://www.conftool.com/tei2019/sessions.php. In a nutshell, this paper proposes an additional eventName element (in parallel to pers|place|orgName) for the purpose of inline referencing, extending the attributes of event with @who and @dur, and further promoting the use of listEvent with the goal of providing a "calendar" webservice that interlinks existing TEI-based digital scholarly editions.

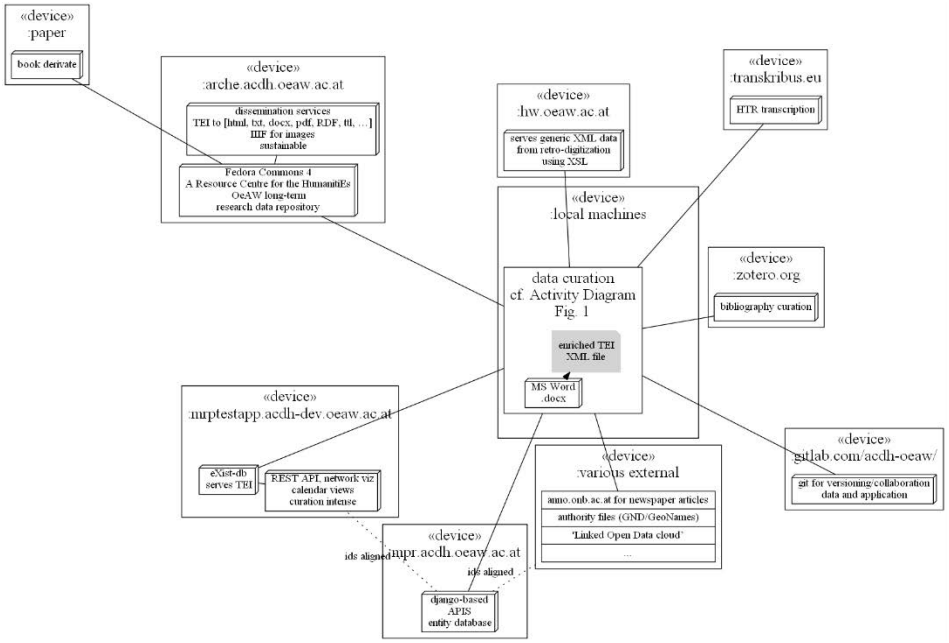

Figure 2. Ministerratsprotokolle 1848–1918: Deployment Diagram. Source: own work

A listEvent of 27 volumes of the Ministerratsprotokolle 1848–1867 edition as outlined above is available at [MRPTESTAPP] under CC-BY licence; it features both the single sessions (2301 entries) and the respective agenda items (10959 entries) which are modelled as events. This is the first time that a complete "table of contents" of all agenda items in the whole first edition series is available to the public (Fig. 3).

The same showcase application also displays a selection of full-text protocol XML files from our current editorial work on three volumes of the 1867–1918 series, which include listEvent data in their teiHeaders.

Sources for the eXist-db application based on the KONDE dsebaseapp are available under [GH01], showcase data are kept in [GH02]. A screenshot of the APIS-based entity database is provided in Fig. 4.

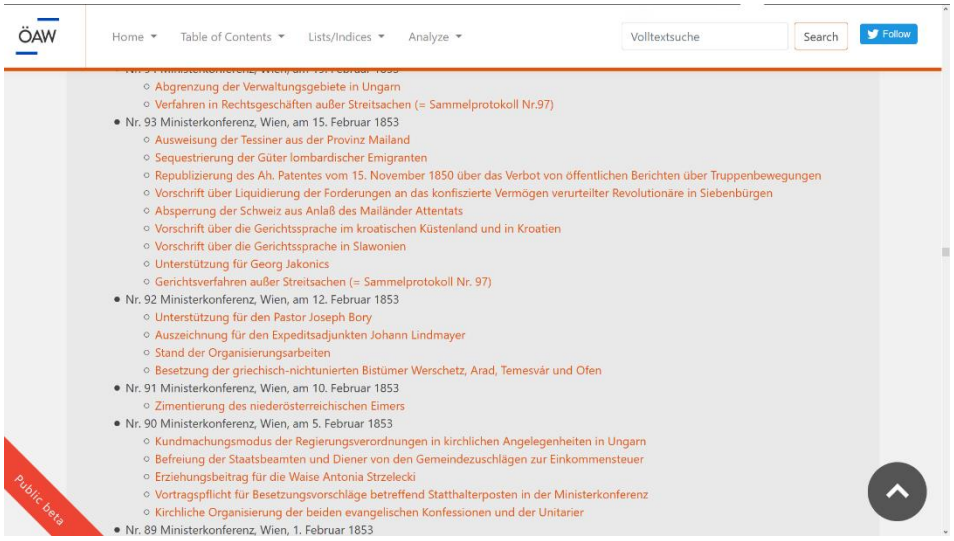

Figure 3. Ministerratsprotokolle 1848–1918: Screenshot of the eXist-db based application [MRPTESTAPP] displaying the 1848–1867 listEvent table of contents. Source: own work

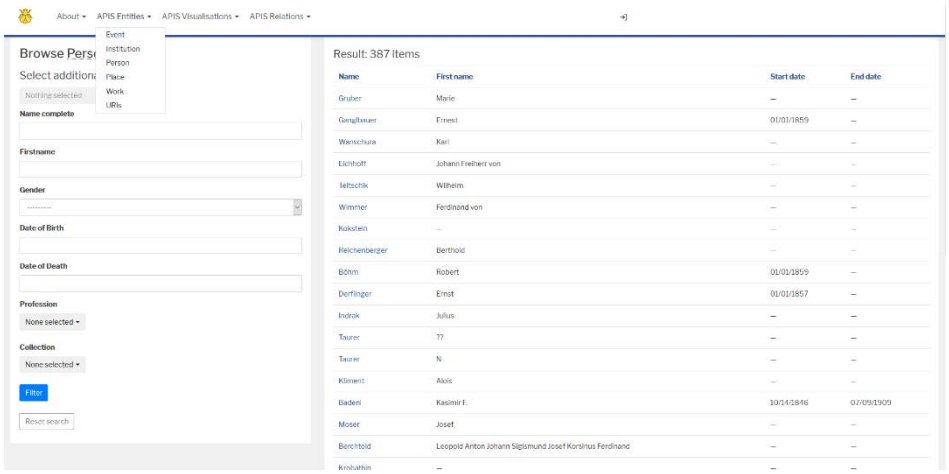

Figure 4. Ministerratsprotokolle 1848–1918: Screenshot of the APIS-run entity database. Source: own work

Conclusion and Outlook

The intention of our project is to open up new access paths to a pre-existing historical textual resource, in a sustainable manner. Much effort has been put into the transcription and commentary of the paperwork that puts the inner workings of post-1848 Habsburg empire governance on historical record. The intention to publish the outcomes of multiple decades of historical research has to include new access routes through non-serial (non-book) representation of the textual data, if we are to follow the digital paradigm that has been articulated e.g. by Sahle 2013, but already by Shillingsburg 2006. In short, this paradigm is defined by the notion that a digital scholarly edition differs from a paper-bound scholarly edition in that it does not merely constitute a digitized version copying the features of a printed book. Instead, it defines a digital edition by its essential incompatibility with a (broadly speaking) linear concept of text-as-book – such an edition cannot be losslessly converted into book form. This includes linking to and/or drawing from external resources like in the case of authority files e.g. with the use case of disambiguation of personal names.

In this sense, the digital edition of the minutes of the Council of Ministers goes beyond the functionalities of the volumes published to date; a print version is nevertheless produced in the chosen hybrid edition approach; it offers a reading typography in a similar fashion to the previous volumes, including all paratexts. In many respects, the future MRP digital edition transcends the paradigm of the book, as it offers:

- the whole set of features from the printed book series, i.e.
    - scientific introduction
    - lists of outdated expressions and participants of the council
    - minute texts, including their abstracts, and addenda
    - indices or persons, places and institutions
- multi-volume facetted full-text search,
- enhanced display and filtering options compared to the print product,
- supplementary facsimiles,[10]
- extensibility in the case of new source finds,
- the addition of authority file data and linked data,
- the development of new audiences through extended visibility,
- also in conjunction with access to the underlying data via API,
- enhanced workflow and data transparency in comparison to previous book production.

Following the claim that “indices of persons, places, and events and calendars and maps are becoming default components for historical digital editions” (Vogeler 2019: 313), our suggestion is to provide a maximum of additional discovery tools with a minimum of additional editorial effort. With listing

[10] As the MRP edition establishes its texts from various different sources, it will not be possible to keep a complete track of source facsimiles within the given time constraints. Still, we publish facsimiles at least of damaged sources (“Brandakten”) with the goal of showing the extent of missing source parts (tei:gap and tei:damage).

almost 11,000 agenda item events with the related governmental staff, a valuable partial data set has been made available already.

Currently, we are not permitted to publish the results of our efforts to standardize and formalize the full-text data of the retro-digitized MRP corpus in TEI markup. Still, we hope to contribute to a discussion on Digital Cultural Heritage that enables researchers not only from the humanities, to make use of the entire corpus that spans over roughly 5 million tokens.

Over the upcoming years, the MRP edition project, one of the long-term projects the Austrian Academy of Sciences has committed itself to, will continue to provide both governmental documents and supporting auxiliary data, thus contributing to historical fundamental research, while also opening up the source data to the public. We are convinced that the administrative history data we provide will spur further research – given the new method and technical dissemination we even hope that it will transgress traditional disciplinary boundaries.

References

Burnard, L. (2019). What is TEI Conformance, and Why Should You Care? // Journal of the Text Encoding Initiative 12, 1-22. http://journals.openedition.org/jtei/1777 (03.07.2019)

Dumont, S. (2015). correspSearch – Connecting Scholarly Editions of Letters. // Journal of the Text Encoding Initiative: Selected Papers from the 2015 TEI Conference 10, 1-21. http://journals.openedition.org/jtei/1742 (03.09.2019)

Grishman, R. (2015). Information Extraction. // The Oxford Handbook of Computational Linguistics. 2nd ed. / Mitkov, R. (ed.). https://doi.org/10.1093/oxfordhb/9780199573691.013.009 (03.09.2019)

Kurz, S., Fischer-Nebmaier, W., Kampkaspar, D., Lein, R., Schmied-Kowarzik, A. (2019). Die Edition der Ministerratsprotokolle 1848–1918 digital: Workflows, Möglichkeiten, Grenzen. // 5. Digital Humanities Austria Konferenz DHA 2018 Conference proceedings / Zeppezauer-Wachauer, K., Hinkelmanns, P., Ernst, M. (eds.). Salzburg/Wien (in print)

Kurz, S., Zaytseva, K. (2019). Herausforderungen für Thementhesauri und Sachregister-Vokabularien zur Erschließung im Kontext des digitalen Editionsprojekts Cisleithanische Ministerratsprotokolle. // DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts / Sahle, Patrick (ed.). Frankfurt/Main

Piotrowski, M. (2012). Natural Language Processing for Historical Texts. // Synthesis Lectures on Human Language Technologies 17. https://doi.org/10.2200/S00436ED1V01Y201207HLT017

Rumpler, H. (1970). Die Protokolle des Österreichischen Ministerrates 1848–1867, Einleitungsband; Ministerrat und Ministerratsprotokolle 1848-1867. Behördengeschichtliche und aktenkundliche Analyse. Wien: Österreichischer Bundesverlag

Sahle, P. (2013). Digitale Editionsformen. Norderstedt:BoD. // Schriften des Instituts für Dokumentologie und Editorik 7-9

Shillingsburg, P. L. (2006). From Gutenberg to Google: electronic representations of literary texts. Cambridge: Cambridge University Press

Parthenos. Standardization Survival Kit. https://ssk.readthedocs.io/ (03.09.2019)

TEI Consortium (2019). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.6.0, July 16, 2019. https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf (03.09.2019)

Vogeler, G. (2019). The ‘assertive edition’. On the consequences of digital methods in scholarly editing for historians. // International Journal of Digital Humanities 1, 2, 309-322. https://doi.org/10.1007/s42803-019-00025-5