CrossMark

A framework for modelling and visualizing the US Constitutional Convention of 1787

Nicholas Cole[1] · Alfie Abdul-Rahman[2] · Grace Mallon[1]

Abstract

This paper describes a new approach to the presentation of records relating to formal negotiations and the texts that they create. It describes the architecture of a model, platform, and web interface (https://www.quillproject.net) that can be used by domain experts to convert the records typical of formal negotiations into a model of decision-making (with minimal training). This model has implications for both research and teaching, by allowing for better qualitative and quantitative analysis of negotiations. The platform emphasizes the reconstruction as closely as possible of the context within which proposals and decisions are made. The usability and benefits of a generic platform are illustrated by a presentation of the records relating to the 1787 Constitutional Convention that wrote the Constitution of the USA.

Keywords Humanities · Data exploration · Negotiated texts · User interfaces

1 Introduction

1.1 Overview of problem

Many of the foundational texts of the modern world have not been written by individuals, but negotiated by groups of people in formal settings. This class of document includes treaties between states, texts such as the Universal Declaration of Human Rights [44], negotiated at the United Nations between 1946 and 1948, innumerable pieces of legislation negotiated in the world's legislative assemblies, and constitutions, such as the one negotiated by the American states in the Constitutional Convention of 1787 [43], which met between May and September of that year. The records of these negotiations are extremely hard to follow. Typically, the journals published after them record the proposals made and votes taken, sometimes with a near-verbatim account of what was said, but often with concise descriptions of debate. Their principal purpose is to help those involved in the discussions keep track of the process of negotiation in which they are immersed; intelligibility for later audiences is only a secondary concern. To fully understand how negotiated texts are created, it is necessary to understand both the temporal sequence of proposals and votes and the hierarchy of decision-making, as proposed amendments are themselves amended and amended before being finally accepted or rejected.

The difficulty of following these records limits their utility for researchers. It also restricts broader use in education and outreach. Even a relatively short document, such as the Constitution of the USA, can be the product of many thousands of formal proposals and votes taken. A further complication is that there is no particular requirement that votes be taken in close temporal proximity to the proposal to which they relate, meaning that the context in which a vote is taken on a proposal can be considerably different to the context in which the proposal was originally made; indeed, the document may have evolved in the interim to the point that even those who proposed a particular change may vote against it. Providing any kind of detailed commentary on this process, or explaining it in classroom settings, therefore becomes extremely difficult. It can take many pages of text to explain the evolution of particular provisions in prose. For example, a ground-breaking

✉ Alfie Abdul-Rahman
  alfie.abdulrahman@kcl.ac.uk

Nicholas Cole
  nicholas.cole@history.ox.ac.uk

Grace Mallon
  grace.mallon@univ.ox.ac.uk

[1] Faculty of History, University of Oxford, George Street, Oxford OX1 2RL, UK

[2] Department of Informatics, King's College London, Bush House, 30 Aldwych, London WC2B 4BG, UK

essay attempting to explain the compromises and manoeu-vres which created the electoral college as the method for choosing the American president had to interweave analy-sis in an attempt to describe the changing circumstances in which proposals were being made and decided [33].

1.2 Particular case study

The Constitution of the USA [43] has special significance as the first example of a constitution for a large state that was negotiated in this collaborative way and a constitution that at the same time incorporated many novel features, not least of which was the compromise between state and national authority that the Convention was able to agree. Those who recommended adoption of the new Constitution were proud of this aspect of its creation. As Alexander Hamilton put it in the first of the *Federalist* essays [41] (p. 301), "It has been frequently remarked that it seems to have been reserved to the people of this country, by their conduct and example, to decide the important question, whether societies of men are really capable or not of establishing good government from reflection and choice, or whether they are forever destined to depend for their political constitutions on accident and force". Although the idea of a collectively written constitu-tion may have been novel, the delegates to that Convention were apparently well familiar with a formal, "parliamentary" style of conducting business. The rules that they adopted for themselves at the start of the process were relatively brief and specify only a few specifics, taking much of the process of debate for granted, a fact which underscores the pre-existing and shared understanding of parliamentary-style processes to those who took part.

The corpus of texts used for this case study consisted of the official journal, other papers preserved as part of the official paperwork of the Convention, and a series of less formal leg-islative diaries kept for a variety of purposes and with varying levels of detail and accuracy by members of the Convention. These records were brought together and integrated by Max Farrand in a 1911 edition of the papers [11], which is out of copyright globally and could be freely used to present an open-access model of the debates of the Convention with-out the need for manuscript work by the editors. We judged that the relatively minor inaccuracies in the Farrand edition would not interfere with its usefulness for the development of our model and the related visualizations. A fuller descrip-tion of the nature of the records used for this project is given by Mary Bilder [2,3]. An online facsimile and full transcrip-tion of the 1911 Farrand edition is available at *The Liberty Fund* [10]. For an overview of the process of debate adopted by the 1787 Convention, with a particular focus on the com-mittee structure, see work by John Vile [46]. These records cover the formal period of debate from May 14, 1787, to September 18, 1787. The records for certain subcommittees do not survive (and it is not clear whether formal records were ever kept), though the formal and informal records cover the work of the plenary sessions and "committee of the whole" in detail.

2 Opportunity for modelling

A full history of parliamentary style, formal negotiations is beyond the scope of this paper. The processes that the British Parliament used to facilitate debate upon and agreement of a legislative text were codified and published in a variety of forms from the early modern period onwards. John Hatsell's *Precedents of Proceedings in the House of Commons, under separate titles with observations* [15] influenced Thomas Jef-ferson's *Manual of Parliamentary Practice* [16]. There have been many publications that seek to capture either the rules and procedures that pertain to one particular legislative body or which offer more generalized rules for use in a variety of settings. What distinguishes parliamentary-style debates from other forms of deliberative assembly, however, is the focus on the production of a specific text, with the specifics of that text not only agreed or rejected, but actually writ-ten (or at least rewritten) by the participants in the process of deliberation. Whereas other deliberative assemblies have historically had the choice to merely accept or reject propos-als, the types of negotiation covered by this paper are those where the final document is the result of work within a formal framework that provides a mechanism for the creation of a multi-author document.

Though parliamentary procedure has always been adapted to specific circumstances and the precise rules of debate will vary from place to place and across time, our survey of the records of parliamentary processes and a range of manuals of parliamentary procedure convinced us that a number of com-mon elements existed in all such processes that enabled the creation of a generic model which could be used to capture the work of any particular assembly. These common features all stem from the fact that for a body of negotiators to agree upon a text, it must at minimum be possible for the participants to know with certainty what has, and has not, been agreed at any particular moment. At minimum, therefore, the core of formal parliamentary procedure revolves around the idea of a text being introduced for debate and opportunities being provided for that text to be amended by the participants. Pro-posed amendments are made in the form of specific changes in wording, which are either agreed or rejected. Since pro-posed changes to a text may themselves by subject to debate, discussion, and amendment that may occur prior to a deci-sion as to whether or not to accept or reject the amendment, keeping track of exactly what has or has not been agreed at any given moment and the current wording of any suggested

amendments is the task of the secretariat that supports any process of negotiation.

In these formal processes, texts are created through a process that consists of discrete and formally defined actions (such as to introduce new documents, to suggest amendments on, to agree or reject proposals) that occur within a temporal sequence and which also that have a hierarchical relationship to each other that can be represented in a tree structure. The records typically left by such proposals record both relationships, though in a format that makes it easier to understand the temporal sequence of events and more difficult to follow the hierarchical relationships, especially during complicated periods of debate. At minimum, a formal journal records the proposals made to amend the text under discussion and the decisions taken upon those proposals, in the order in which they happened. This is the minimum record-keeping required for the actors in a process of negotiation to be sure of the text that they have agreed. Thus, the official journal of the 1787 Constitutional Convention, which was not intended for publication, does not record speeches made on either side of a question or even (in many cases) the precise division of votes [11]. It records instead the minimum set of information necessary to allow members of the Convention to keep track of what had been formally agreed and to settle any disputes about the wording that had been negotiated.

Since the nineteenth century, however, it has been much more common for the records of legislative assemblies and other formal negotiations to publish the records of their work, and the trend has been towards the publication of verbatim records of deliberations, in part to memorialize the participation of the actors involved. An exhaustive list of such journals would be impossible. In the USA alone, all state-level legislative bodies have published official journals, as have the two chambers of the Federal legislature. There have been more than 210 constitutional conventions for the purpose of writing a state constitution in the USA since 1776, most of which have published the records of their deliberations. In the field of public international law, the negotiation of (for example) the Paris Peace Treaties of 1919 and the foundational documents of the United Nations have all been accomplished through parliamentary-style processes of negotiation and have resulted in formal records of the deliberations.

Yet whether or not they are verbatim transcripts, a record merely of proposals made and decisions taken, or summary minutes that fall somewhere between the two extremes, the formal records of such processes are hard to follow. A participant within any such process would have had access to something which is for the most part lost for any subsequent reader: immediate understanding of the current state of any document they were discussing. Indeed, it is the task of the secretariat supporting the negotiation to provide them with this information.

The insight of this project is that the very formality of these processes and the nature of the typical records make it possible to reconstruct the formal context—that is, the agreed state of the various texts under consideration—within which events happen, and that doing so transforms the utility of this class of records for a variety of users. Furthermore, the creation of a standard model which can capture work of such deliberative bodies enables meaningful comparison to be made more easily between different processes of negotiation.

In this paper, we present the following contributions:

– We have created a generic platform which can be used to encode the records relating to a wide variety of negotiations and which can stand alongside and integrate with existing presentations of those records.
– We have created a generic model for representing the negotiated of texts proposed by multiple actors.
– We have created a platform where each proposal to change text needs to be entered only once, even if the context in which it is voted on is significantly different to the one in which it had been created.
– We have established the conventions for the use of this model in a consistent fashion.
– We have developed visualizations to assist a range of users with varying levels of expertise to explore this material.
– We have developed a web-based application to enable domain experts to construct the model of a negotiation with minimal technical training.
– We have evaluated the ease with which this platform can be used by both domain experts and non-expert readers.

3 Related work

The platform presented here can be distinguished from three broad approaches to the presentation of this class of record. The first is a digital presentation of archival and published records, whether as images of surviving documents or as transcriptions. The second is an attempt to extract (often on a large scale and with as much automated processing as possible) and visualize certain quantifiable aspects of legislative debates preserved in official records. The third approach has been to use visualization and other tools to identify the logical points of contention within a debate—often with the aim of influencing and improving the quality of discussion.

3.1 Image capture of original manuscripts and detailed transcription

Parliamentary records are typically presented online as a photographic representation or transcription of original

manuscript or print journals [8,9,21–23,40]. A number of these projects have worked to provide additional and more consistent structured information to the records as part of the transcription process, so that information about (for example) specific speakers captured in the records can be more easily extracted from the database, or to standardize markup so that transcriptions can be better analysed or connected through the "semantic web" [1,46]. While these provide wider access and may, if a transcription is available, be more easily searchable, the problem for the reader of fully understanding the context of proposals and decisions remains. On the websites of the UK Parliament and other modern legislative assemblies [30,42], graphics are sometimes provided to illustrate the flow of a bill through the stages of being referred between committees and chambers, but these provide only a very high-level overview of the process. These projects focus on providing access to the records of negotiations. Unlike our project, they do not focus on modelling the negotiations or helping users to understand and analyse the records presented. The work presented here is not a new way of presenting the transcriptions of official journals. Rather than duplicate those efforts, we have linked the objects in our model to the transcribed records provided by others, where available. The work presented here seeks to make the records of parliamentary procedure more intelligible, and our focus has been on the development of a navigable model of negotiations, rather than on the transcription of the records themselves.

3.2 Visualization of extracted data from large corpora

Other efforts to provide visualizations of the parliamentary process have focused much more on voting behaviours or other quantifiable data. Some of the most comprehensive visualizations are provided by the *Legislative Explorer* project [6]. The project provides various animated displays that show the movement of bills between the various committees of the US Congress between 2015 and 2017, tagged by sponsor and keyword and enabling a variety of plots to be drawn. Such visualizations, however, are not concerned with the evolution of the text of the legislation in question, but rather on visualizing quantitative data that can be extracted from the dataset, such as the number of bills that have been sent for discussion in particular committees of congress, the number of bills that have become law, number of bills introduced to congress that have been tagged with particular keywords, or which have been introduced by particular members of congress.

Another project that works on the visualizations of the parliamentary process is *La Fabrique de la Loi* [18]. The research question at the heart of this project is the extent to which legislation introduced by the French executive branch is modified by members of the French legislature during its passage through Parliament. Using the extremely formal and predictable structure of the records published by the various branches of the French government (including a clear and detailed combination of reference numbers and headings) and other non-profit groups, this website is able to answer this question by examining the extent to which the text of French legislation has been altered by each the eleven formal stages through which French legislation passes before becoming law. Web scraping techniques are used to capture the text of French legislation at the end of each formal stage (when a full text of the document is made available to the public) and to link those texts to the records published relating to both amendments offered for debate and the debates on the text of the legislation themselves. The way in which French legislation and the associated legislative debates are published means that it is possible for this website to relate the records of amendments to particular sections of a given bill and to display an indication of both the political party responsible for a particular amendment and whether the proposed amendment was adopted or rejected.

This allows the project to answer its primary research question—how much legislation is altered by the parliamentary process?—and also to provide a quantitative answer to a secondary set of questions: Which parties in the French legislature are most active and have the most influence when particular pieces of legislation are debated? The techniques used by this project, therefore, allow aspects of the parliamentary process to be better understood, and the web scraping techniques allow the project to process many years' worth of data. However, the project relies upon the particular nature of the current French record and is in that sense difficult to generalize. In addition, the exact context within which a particular proposal was debated is not reconstructed. The project offers excellent summary information that allows it to answer high-level questions about the work of the French legislature, and excellent links to both official and unofficial records, but it does not allow users to understand fully the context of decision-making. The platform presented in this paper differs from that approach in using the records of a negotiation to offer more detailed reconstruction of context, which entails much more human interpretation of the records in question in order to build its model of the debates. It is therefore a more flexible platform that could be used to model a wider range of records, but one that could not rely on an automated interpretation of the records available.

3.3 Interpreting the logic of a discussion

Computers can be used in many ways to help us understand and improve the way that arguments and decision-making are structured and take place. Unlike the work of Kirschner et al. [17], Reed et al. [31,32], Shillingsburg [36], Winograd

and Floresour [48], and Walton [47], or a commercial platform such as *SEA System* [39], our platform seeks to represent parliamentary processes as they have historically occurred; it does not seek to use computers to impose a different and better structure on the process of debate and argument. It is focused on understanding the creation of documents via historical (or contemporary) parliamentary procedure, rather than on visualizing, refining, and influencing the structure of arguments. Our platform focuses on reflecting the timeline and hierarchy imposed by the formal rules of debate in parliamentary settings, which may not strictly reflect the logical structure of arguments.

3.4 Methodologies from other fields

Formal negotiations are not the only circumstances in which a crucial problem is to track changes to a set of texts. Before designing the platform presented here, the authors examined approaches taken in other applications. The most recent versions of Text Encoding Initiative (TEI) have introduced the concept of "genetic editions" [5] that can be used to describe changes to manuscripts. The potential of this genetic edition approach is shown by projects such as *Digital Variants* [7]. *Digital Variants* presents the variations between authors' manuscripts, drafts, and published editions, highlighting variations in texts. In each case, however, they are dealing with a relatively small number of texts to compare; by contrast, a process of negotiation needs to be understood as (even in our case study) thousands of variant texts. None of the various XML transformation languages that we evaluated seemed appropriate for the specific task we had identified. In particular, they are too sensitive to changes in a document that make it impossible to describe a transformation once and apply it in evolving contexts in a way that would not produce undesired results, meaning that a proposal would have to be encoded multiple times for different contexts, and we judged that using this technology to describe the evolution of text in a formal setting with hundreds or thousands of changes would be cumbersome and error-prone.

The distributed version control systems [13,26,38] used to track and reconcile changes to computer source code during software development solve many similar problems to the ones presented here. Multiple authors are working on changes to a project that are ultimately reconciled to provide a single, agreed set of files. Most of these systems, however, are tightly tied to the specific work-flows of developers and the line-oriented texts with which they are dealing. Some projects have attempted to use Git directly to model parliamentary-style processes or the state of legislation. For example, the *Bundes-Git* [4] project attempts to track the current state of German Federal law. While it is technically possible to use Git [38] to model the process of amendment of legislation, doing so requires working within the limitations of a platform fundamentally designed to store the development of software code rather than the workings of a deliberative assembly. For example, decisions to adopt or reject amendments must be stored as the merging or closing of branches, while no mechanism easily exists to track the membership of particular bodies. In a similar way, a number of platforms, such as those offered by Wikipedia, Google Docs, or Apple's iCloud services, offer the ability for a document to be collaboratively edited, and for the version history of documents to be viewed. However, the model we present here much more naturally represents the workings and complexity of a parliamentary-style process.

Our approach was to adopt the diff–match–patch algorithm [12] developed and released by Google as part of its Google Wave project (an implementation of Myers's algorithm [28], coupled with a mechanism for applying "fuzzy" or "inexact" patches to a base text). It provided a more promising starting point for a platform concerned with the better presentation of the process of negotiation. Unlike many algorithms for describing the changes to documents as patches and applying those patches, the Google diff–match–patch tools were designed to be used in an environment where multiple authors might be working on a document at once, and the order in which patches were received by the participants might vary. This is analogous to a situation in which a proposal may or may not be incorporated into a document, depending upon whether a decision has been taken on it, and in which the changes that a particular proposal would make to a document might have to be made to a different base text to the one in which it was suggested, depending upon the decisions that have been made in the meantime. Working through a variety of test cases proved that this implementation could be configured and used to track the process of negotiating a document.

4 Application background

The records related to the Constitutional Convention of 1787 include an official journal (kept by the Convention's secretary, William Jackson), and a variety of private diaries, the most famous of which was kept by James Madison. The official journal was entrusted to George Washington along with various related papers and published by congress in 1819 [2]. Madison's journal was sold to congress after his death and published in various editions. In 1911, Max Farrand published *The Records of the Federal Convention* [11], a compilation of the various extant records, arranged to allow the parts of the various accounts relating to each day to be read alongside each other. Thomas Jefferson's *Manual of Parliamentary Practice* [24,45], published in 1801, provides a more detailed explanation of parliamentary (i.e. formal negotiating) procedure as it was understood at the end of

the eighteenth century. We compared this to other, similar manuals [27,34], and produced a model of formal negotiation produced as a result. Although the intricacies of the rules vary significantly between legislative bodies, we envisaged a platform which modelled negotiations, not one which enforced particular restrictions. We therefore constructed a model which could be used to model the creation and negotiation of text, and the passing either of draft documents or amendments between committees, as well as one which could track committee memberships across time. We constructed a series of test cases based on the analysis of these parliamentary manuals, to ensure that our platform would be able to model any likely action by a legislative assembly, recognizing that in practice legislative assemblies frequently suspend their own rules or behave in surprising (and less than entirely logical) ways.

Although there were certain gaps in the records kept of the Convention (principally, the work of smaller committees), the records that did exist seemed to contain enough details of formal proposals and votes taken that the process by which the US Constitution had been negotiated could be reconstructed in large part. Certainly, for better-documented processes of the same type, the official journals provide enough information to reconstruct every step of debate. Jackson's 1787 journal did make an effort to record both the wording of formal motions and the outcome of votes, and covered the work of the plenary sessions and the work of the Committee of the Whole. What is often less clearly recorded or even absent is the record of which way the various delegations voted on any question. In addition, the precise sequence of events within a given day was sometimes recorded differently by the official journal and the various private journals [11]. What was not known with certainty even to specialists before we began this project was whether the origin of every single clause of the final text could be accounted for by the extant records. We believed that a platform that could satisfactorily model these particular records would have broad application, since in many other processes the kind of uncertainty presented by these sources does not exist.

5 Design notes

Our primary requirement was the creation of a platform that could present the state of documents during any moment of negotiations. This would involve storing a representation of the sequence of events within a negotiation (the linear timeline of each committee), in such a way that the agreed state of documents and related information could be calculated and presented to users for any selected moment in time. We anticipated that the research assistants employed to enter the data would need to have excellent historical skills (because of the issues with the source material outlined above), but with little or no programming experience. We wished to have an interface for data entry that would be intuitive for users without much technical training, and which would encourage the model to be used in consistent ways over the course of a long project.

We did not wish to duplicate the efforts of other projects. In particular, the images of manuscripts and historic printed sources, transcriptions of those documents, and biographies of those in the 1787 Convention have all been presented online by projects at public institutions such as *the National Archives* [29] and *the Library of Congress* [20], non-profit organizations such as *ConSource* [37] and *the Liberty Fund* [19], and by projects based in universities, such as the *Electronic Enlightenment Project* [25]. All of these projects offer bespoke tools based on their specific expertise and the nature of the material they are presenting. Rather than compete with their efforts, we decided to make it possible to associate links with these resources with specific objects within our database, and also provide methods for other projects to link to related information within our own platform.

We knew that those entering data to our platform based on the interpretation of primary source data would need to exercise a certain amount of judgment in interpreting the sources and that mistakes in data entry were possible. The source materials would raise issues that needed discussion among the editorial team, and the decisions taken would need to be documented. Since this project would involve building a model based on the source material, rather than a more mechanical process of transcription, the verification of data entry would involve human judgment. We would need the system to be able to show those running the project which sections of the data entered into the platform had been checked and by whom.

Since the purpose of the platform was in part to allow detailed commentary to be more easily written on the process of negotiation, we envisaged a system of "commentary collections" that would be owned by single or multiple authors. These collections would consist of an introductory essay and then explanatory text linked to specific events within the platform. These collections were to have two functions. First, they could be presented to users who were viewing a section of the timeline to which they were relevant. Second, they would provide an alternative way to navigate the timelines that the platform would present. Users following the debate over particular topics would be able to use the commentaries related to those issues as a guide.

We decided that the main user interface for both users and editors would be built to work within a web browser, while the database and processing would be stored on a central server. Offering a web-based application would be attractive to a broad range of users without the need to install special software, either for viewing or editing, and would improve

Fig. 1 Four categories of event used in the Quill platform model of a formal debate
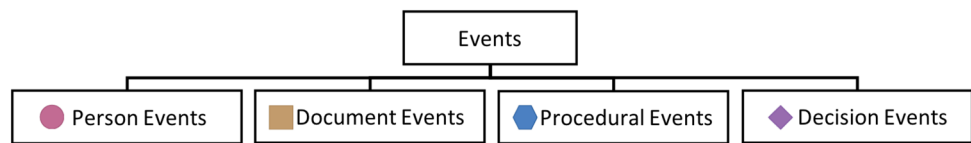
our ability to collaborate with teams working at other institutions. We would incorporate a flexible permissions system within the platform so that different categories of user could be given specific permission to view, edit, or verify specific information within the platform. However, once material had been checked and approved for publication, we wanted as much of the platform as possible to be usable without registration. We also want the web-based interface to be highly flexible from a methodological perspective.

Due to the imperfect nature of the records from which we were working, we knew that our platform would also have to be able to capture the variation between manuscripts and, to a certain extent, uncertainty as to what had actually happened at particular moments. This was especially true of the records of particular votes, where it was frequently impossible to be certain who had voted in particular ways, even if the outcome of the vote was known.

We built the platform around several clearly defined categories of user formally defined within the system, and with several broader communities in mind. For the purpose of building particular models, users of the platform could be designated "senior editors", "editors", or "contributors", with a variety of associated permissions to enter data into the platform, edit the work of others, mark data as reviewed and correct, or approve data for publication. Users can be given entirely separate permissions to be "editors" or "contributors" to collections of commentary annotations or other resources provided to users of the platform, and given permission to associate these collections with items from specific models. We were keen that all data entered into the platform would be clearly attributable to specific individuals, both for the purpose of auditing the information presented in the platform and to encourage the best scholarly practices. Many projects within the platform are private, though the 1787 model and many associated resource and commentary collections are publicly accessible. More generally, the platform is intended to be of use to several distinct communities of users. Those working to model particular negotiations within the platform are one, and researchers seeking to use these models either to investigate particular processes of negotiation or else to compare one process to another are another. We also aim for the platform to be of utility for more general readers and in classroom settings, and are aware that such users may require "distant" readings and more guided routes through the platform more often than close readings of individual moments of a negotiation. To date, our efforts have prioritized building the correct tools for the first group.

These are the tools that we have most extensively tested and evaluated. We have made good progress in the development of tools for advanced researchers and have a roadmap of further tools for detailed analysis and comparison that are in development. We continue to work with partners in the USA to better understand the needs of the third group of users.

6 Building a model of a negotiation

After an analysis of the common features of parliamentary-style processes and our requirements, we created a data model that captures the discrete and important elements of a process of negotiation. Each process, or "Convention", consists of two main components drawn from the historical data (the "Delegation" and "Committee" objects) and two that enable us to connect our model to other data ("Resource Collections" and "Commentary Collections"). Each Delegation is a collection of the "Person" objects that comprise it, while each Committee stores details of its "Sessions" as an ordered list, and within each of those the details of the discrete "Events" that take place within that session, again as an ordered list. The most important part of the model are these Event objects, which record the business of negotiation. The Event types were further broken down into four categories (see Fig. 1), which are sufficient to capture the work of all formal negotiations we have so far tried to model, and which allow us to capture the procedures outlined in the various handbooks of parliamentary practice we have surveyed. Firstly, those that concern the creation and proposed amendment documents. Secondly, those that concern the role of individuals (principally, membership of particular committees). Thirdly, those that concern proposals that do not directly (but might indirectly) affect the creation or amendment of documents, such as motions to adjourn, or to rule particular proposals out of order. Lastly, those that concern decisions taken. "Voting Records", storing the details of particular votes, are linked to relevant Event objects.

Consistent use of the model presented by the platform was ensured in three ways. First, a policy document was kept by the editors, and updated as specific issues were encountered. Second, the different types of events tracked by the platform were kept to a minimum to allow for an accurate representation of the parliamentary process. In general, we found that users with a small amount of experience with the platform would use the model in consistent ways because the platform

presented them with obvious choices in most situations and required them to make relatively few decisions about how the model would be used. Two to three days of training have proved sufficient with a variety of advanced undergraduate and graduate students. Third, we designed the user interface to force users to capture the sources from which they were working in standard ways, usually by automatically validating the input and requiring active choices within the dialogue boxes presented to users, rather than offering default selections. We had to balance this with ease of use and the likelihood of error, and made adjustments where appropriate during the four months of the data entry phase of development in 2016 on the basis of feedback from those doing the work of data capture and verification.

7 Event processing

The architecture of the *Quill* platform is presented in Fig. 2. Objects stored in the database fall into two main categories. One sort of object exists in strict hierarchy of relationships that allows both the sequence of events within a negotiation and their procedural relationship to each other to be stored. An event processor is used to extract meaningful information from this hierarchy. A separate category of data is not accessed through the Event processor layer, but rather stores

Fig. 2 *Quill* platform architecture

information that is meaningful outside the context of a particular moment in the Convention. Much of the metadata that describe a given event fall into the latter category, and there are significant advantages to maintaining the separation. For example, though the fact of a decision (whether a given proposal was adopted or rejected) is intrinsic to the nature of that event, the record of who did and did not support the decision could be (and, given the nature of the extant records for 1787, often is) the subject of conjecture or controversy. Most of the logic to interpret and reconstruct the "timeline" events, therefore, exists in the event processing layer rather than the database itself. The web interface combines information from both the timeline-focused and static information database in the tools that it presents to the user.

Events themselves were subdivided into four main groups:

– Those related to the creation and editing of documents (including the creation of documents, proposals to amend them, or the point at which a document or amendment was passed for review from one committee to another).
– Events related to people (when an individual joined or left a committee, for example, the election of an individual to a particular office, or a roll-call of who was present at a particular moment).
– Events related to the "procedure" of a given committee, such as a motion to adjourn, or more complicated motions that have the effect of invalidating or altering previous decisions.
– Events that record a decision that has been taken (whether implicitly or explicitly).

Within this system, documents are represented as a proposal to create them and a series of proposals to amend them, together with the decisions taken on those proposals. An event processing layer is able to reconstruct the state of the documents and committee memberships for any given moment in the timeline reconstructed from this database. It is this event processing layer that contains the algorithms for merging together documents on the basis of proposals made and votes taken. This algorithm must account for the fact that the state of a document relevant to a particular moment of debate must also take into account the proposal under discussion. That is to say, that a proposal to change an amendment that has not yet been accepted needs to take into account its parent amendments and any sibling amendments that have been agreed, but not any siblings that have not been agreed, nor other pending proposals that have not yet been resolved.

For any given moment of the timeline, therefore, the processing layer is able to calculate:

– A list of documents currently under discussion.
– A list of proposals that have not yet been resolved (the "pending" proposals).
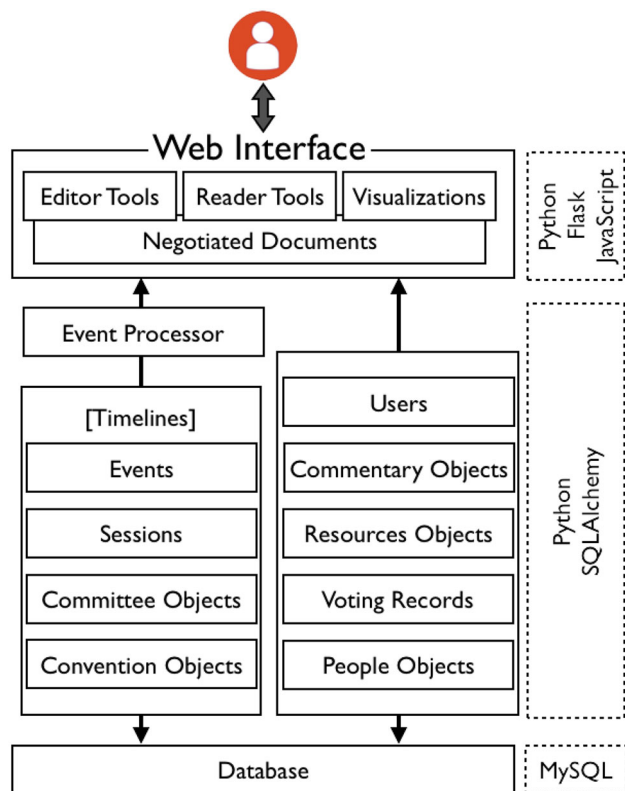
– The state of any documents currently agreed or under discussion.
– The state that those documents would be in if any of the pending proposals were adopted (if it is possible to generate this).
– The current membership of the committee.
– The information necessary to display various visualizations related to the document and proposal under discussion.

This can be combined with other information, such as links to further resources, that are associated with that moment in the timeline.

There is a potential for confusion to arise over the use of the term "timeline". The data model that we propose here focuses on the elements of a process that result in the transformation of the texts under discussion. Each "event" is therefore a discrete action that results in the transformation of one or (in certain special cases) of a number of texts. Our underlying model does allow for the exact timing of individual events to be captured; however, this is not currently exposed in our public interfaces, for the simple reason that in almost no instance do the published journals of a negotiation record the timing of events with that level of detail. Indeed, since the records we were using for the case study presented here were not even verbatim transcripts of the proceedings several hours of speech-making might be recorded in a single line of summary (or not even recorded at all!). The "timelines" presented throughout our platform record the sequence of events, but each event is a discrete action that affects the agreed texts of the documents under discussion by a negotiation. A speech that is many hours long might be captured in our platform by a single "event", while at periods during which complicated changes are being debated, several "events" in our timeline might all take place within the space of a few minutes.

7.1 Text processing

The text processing layer of our platform implements an algorithm that takes a series of events and calculates the set of proposals that should be included to create the various versions of a text relevant to a particular moment in time. The formal text of the proposals themselves is stored as a series of transformations encoded as *diff* objects. Our algorithm produces texts by combining applying these proposals in the order most likely to result in the intended texts and makes the necessary adjustments to the sensitivity of the match and patch algorithm to allow the document to be built.

Our implementation uses the diff–match–patch algorithm [12] and work on plain text. This imposes a set of restrictions on the nature of the data we can store. In practice, we have not found this to be limiting. Processes of negotiation from the period when parliamentary-style procedures were first formalized (in the sixteenth century) until the widespread availability of word processing technology are well captured within our system. Even though word processing has allowed for more elaborately formatted documents, this has in practice little changed the character of the formal documents considered within a parliamentary process, and we are currently successfully modelling the passage of the Brexit legislation through the UK Parliament, albeit with a small number of compromises as far as the presentation of documents is concerned. However, our architecture has been designed to allow the algorithm used to merge together proposals to take advantage of developments in diff–match–patch-style algorithms, and we are currently working to adopt an alternative approach that should allow us to process more structured documents and to model modern, word-processed documents without compromise and within standard markup frameworks, at the cost of significantly increased complexity both in processing and in the data entry interface.

8 Visual interface

We created a web application called *Quill* (https://www.quillproject.net) that would be used both for data capture and by readers. We were aware that readers would not all require the same level of detail. Whereas those working on data capture would be most concerned with the detailed sequence of events within a committee session, many users of the platform would be better served with a more general overview of events, and we are also conscious that the visualization needs of advanced researchers differ from those who wish to use the records in classroom settings or for general interest.

For the bulk of data entry, which concerns the creation of the model timeline, a view of the committee session in question is presented (see Fig. 3). The timeline of the committee is represented by a horizontal series of icons. Around this is displayed information useful to those translating the source material into the model. The current membership of the committee is shown, together with summary lists of proposals that are pending for debate. Between each event of the timeline, users are offered a ⬚ button to allow them to see the exact state of any documents or proposals at that point, and a ✚ button that allows them to add a new event to the timeline at that location. The latter causes a pop-up form to appear, in which the user is invited to select the type of event he/she wishes to add (see Fig. 3)—for example, a person joining a committee, or a proposal to amend a document. The fields of the rest of the form are adjusted based on this choice. This session display also includes tools for editing existing events and deleting them from the timeline, and other functions needed by those entering data or verifying data entry.

Most of the other information required by the platform is entered on the "Full Record" view. This is where the

Fig. 3 "Full Record" view of the *Quill* platform showing the session view interface for data entry and editing

names and members of particular delegations are stored and where the names of different committees are created. A page devoted to each committee shows a listing of all of its sessions and allows those to be added to. This page also shows users with the appropriate permissions an overview of which committee sessions have been verified and by which users.

New users of our platform are guided to the "Secretary's Desk" view (see Fig. 4). This combines a list of all of the Convention's Subcommittees, a sense of when they met (represented simply as a timeline showing their first and last session), and a timeline of the individual sessions for any selected committee. A small chart under each session gives a quick sense of the number of individual events contained within it, while mouseover information provides a more detailed view. For any selected committee session, we display a list of documents under discussion, indicating any whose text is altered by that day of debate and any unresolved proposals that relate to a selected document. If users select a document, its current text is displayed in the centre of the screen, and users can choose to highlight the text that was altered by the selected session's debate. If they select an unresolved proposal, they can likewise see the effect that adopting that particular proposal would have on the state of the document. If users want more details, they can easily click
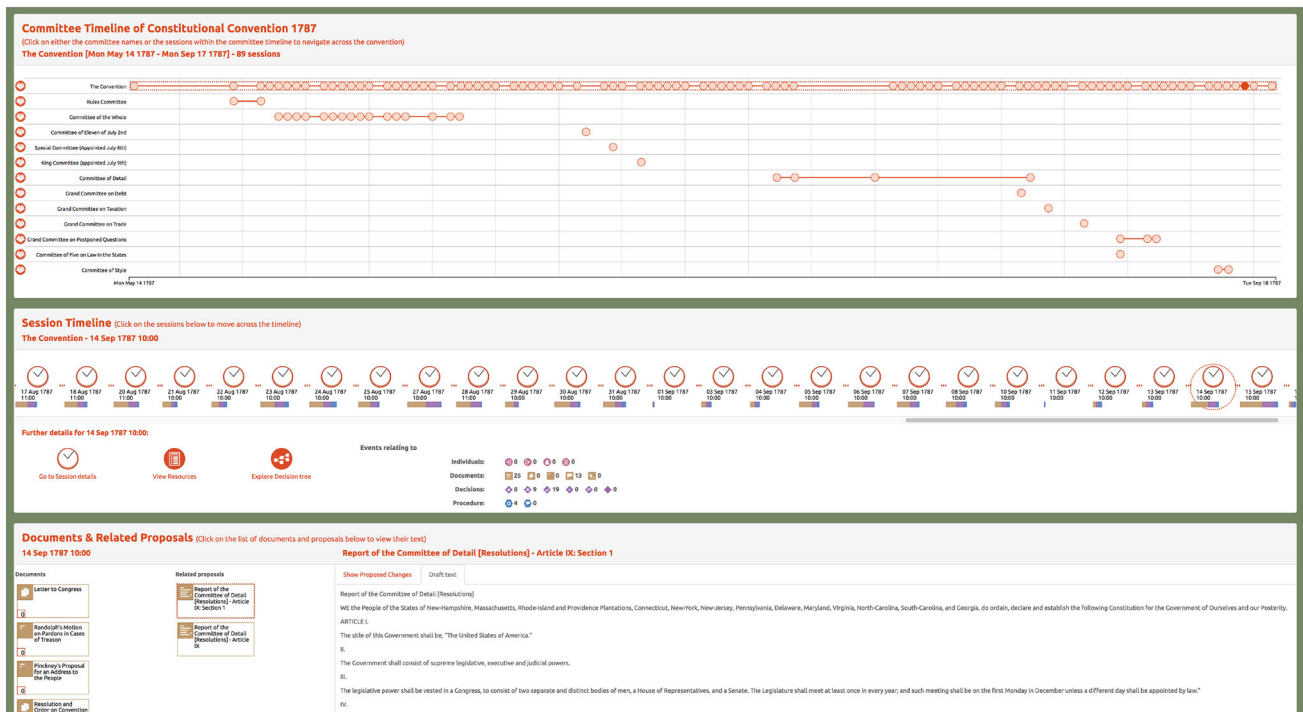
Fig. 4 "Secretary's Desk" view, used to introduce users to the basic functions of the platform

through to the more detailed session view. Though this introduces users to the concepts used in other parts of the platform and provides an effective overview of the process of negotiation, it is frequently a misleading display, since it will often be the case that in practice committees will work for a whole session on changes in wording that they ultimately reject.

However, a similar set of metaphors is maintained in the more detailed visualization of each session (see Fig. 5). Along the top of the screen is a horizontal representation of the timeline of that committee session (see Fig. 5b). Down the left-hand side of the screen are lists of documents and proposals currently under discussion (see Fig. 5c). Users who click on any of these documents are presented with their current agreed state, and clicking on any of the pending proposals shows the state of the documents that they would create. The centre of the screen contains the text relevant to the proposal selected in the timeline. For proposals to amend documents, this is:

– The "agreed text" of the document. This is the text of the document if the document were simply accepted as final in its current state, with no further debate.
– The "proposed text" envisaged by a particular proposal.
– The "intermediate text" that this proposal amends. That is to say, the state of the document including any parent proposals of this text including any relevant sibling proposals.

– A display (the "markup text") that highlights the difference between the "intermediate" and "proposed" texts.

On the right-hand side of the screen is an area where users can choose to display either the details of the selected event or a variety of other tools. A "Document Complexity Tree" shows all of the proposals and decisions that have formed the document relevant to the selected proposal into its current state (see Fig. 6). The proposal under discussion appears highlighted at the top. By navigating both the linear timeline at the top of the screen and using the complexity tree presented on the right-hand side of it, users can quickly understand the relationship between different proposals and the way in which they shape the creation of documents. The display of the tree of decisions that make up a document shows the extremely careful and often word-by-word nature of these negotiations, and provide an alternative method of navigation, allowing users to navigate decisions by hierarchy rather than by timeline. As other negotiations are modelled, it will become possible to compare the structure of decision-making between processes. Further tabs provide access to commentary collections relevant to this event or links to resources held in other collections relevant to this event.

A particular challenge throughout the platform has been to capture for users the relationship between the sequence of events and the formal hierarchy that relates proposals to one another. In earlier iterations of the platform, these were presented separately, but in more recent versions we have
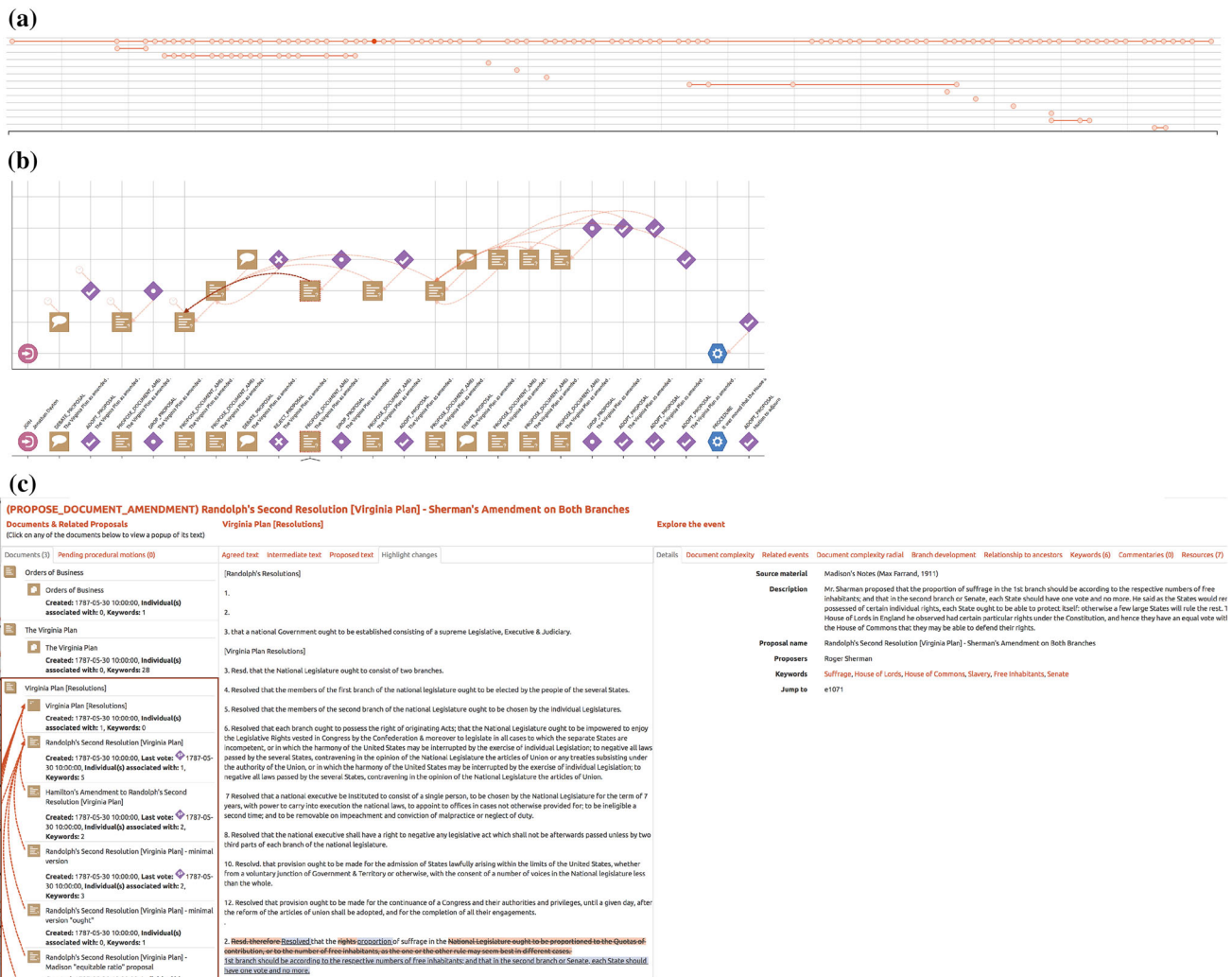
(a)

(b)

(c)

Fig. 5 Presenting the session to users. **a** A map of the Convention to help users orientate themselves within the process by showing them where they are in the Convention, **b** two ways to present the timeline, showing both sequence and hierarchy (top) or just sequence (bottom), **c** part of the visualization of a session intended for readers, showing information relevant to a particular moment

displayed the relationship of events on a two-dimensional display (see Fig. 5b). The sequence of events is represented by the *x*-axis, but the hierarchy of events is represented by placing them differently on the *y*-axis. Proposals to create documents are placed at the bottom of the display, and proposals to amend them or votes taken are placed at higher positions on the axis. Thus, proposals to amend the root document appear at one level, while amendments proposing changes to other amendments are clearly visually separated. The relationships are further clarified by drawing arcs to show the relationships between events. If an amendment relates to an event in an earlier session, a small icon is displayed alerting the user to that fact. Once users learn to interpret these displays, they convey quickly the character of the negotiation as it changes over time. For example, the display for the Constitutional Convention on September 15, 1787,

shows a large number of changes to the root document itself, reflecting small changes in wording to a document that was nearly complete and where most of the controversial issues had been well worked through. By contrast, earlier stages of the debate have shorter displays (fewer distinct changes are proposed and decided upon) but of greater depth, reflecting dispute over controversial issues and competing proposals or forms of words being presented for consideration on particular topics. On the July 11, 1787, for example, the work of the Convention turned to the question of how changes in population within the USA should be measured and who should control changes to the form of the national legislature as a result. Users who are familiar with these displays quickly learn to interpret them and find them a powerful tool to assist navigation and interpretation. In the current version of the platform we use versions of these displays in the individ-
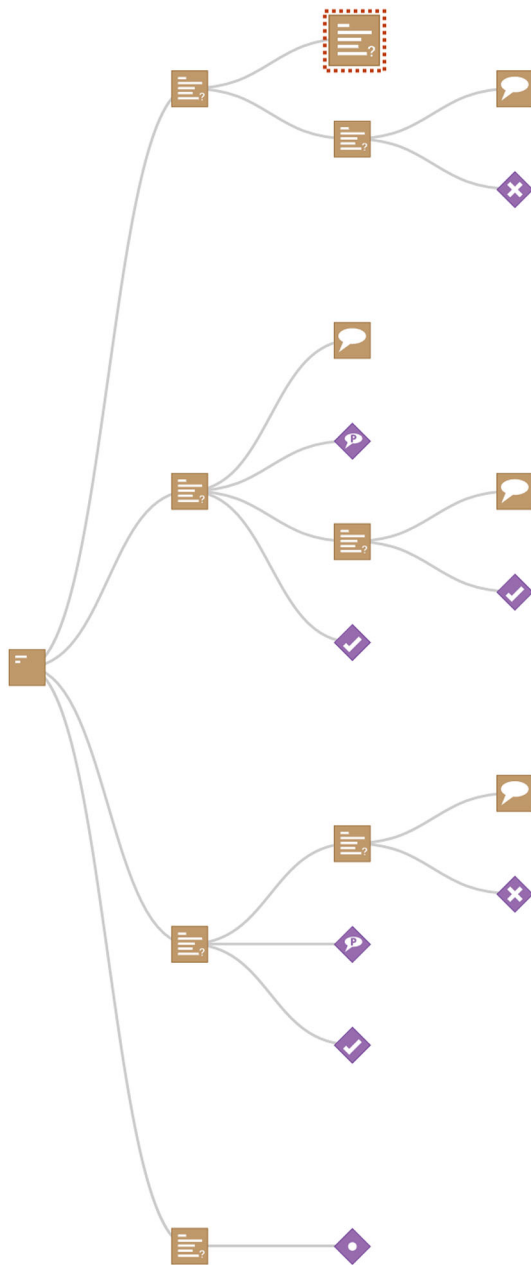
Fig. 6 An example document complexity display from early in the Constitutional Convention. This shows a root document with four amendments, all but one of which were further amended before being accepted or rejected, and the decision taken on each proposal. This view represents only the formal relationship between proposals, and not the sequence in which they occurred, but allows the user to identify controversial areas of debate, as well as providing a representation of the detailed work required to agree even a short piece of text

ual session visualizations, but retain the more familiar and instantly intuitive single-axis presentation of the timeline as well, allowing users to use the two-dimensional version as they become more comfortable with the platform.

Readers might want more of a sense of the structure of decision-making within the Convention as a whole. We can display the flow of documents between committees, or a display representing the overall hierarchy of decision-making within a process of negotiation, presented as a radial tree with the various committees, the documents being considered, amendments on those documents, and any subordinate amendments or decisions, radiating out from the centre (see Fig. 7). A particular challenge is to fit this display on to smaller screens, and we offer users a choice of compact and expanded views. A search box allows users to highlight events on this display based on the metadata associated with each event. It is useful as a display which provides new users with a sense of the complexity of the work of a formal process of negotiation, even though it shows only a hierarchical and not a temporal relationship between events. We have observed that new users find it an inviting way to begin exploring the work of the Convention in a nonlinear fashion and helps them to gain the confidence needed to explore more detailed tools.

Another challenge was to allow users to examine the detailed work of the Convention while maintaining a sense of the overall shape of negotiations. Originally, the displays of the work of each session simply showed the date of that session, but experience proved that it was easy for users to become disorientated, especially if using any of the tools (such as search tools or commentary collection links) that allow them to navigate the material in nonlinear ways. After experimenting with various options, we developed a representation of the negotiation process as a whole that shows the work of each of the committees on separate timelines, with a dot showing when the each meeting of a committee took place (see Fig. 5a). This has proved to be an effective aid to navigation, and we have used it throughout the platform where appropriate, both to help orientate users when they are looking at the more detailed displays in the platform and also to give a sense of the areas covered by a particular commentary collection.

The influence of particular delegations within the Convention is captured on a summary screen that shows two graphs. The first of these is created from a principle component analysis of the matrix of votes within the Convention and gives a sense of how likely different delegations were to ally with each other. Users can choose a one-, two- (the default), or three-axis display. A bar graph represents how many proposals were made by members of a particular delegation and how many of those proposals were accepted or rejected (see Fig. 8). This captures a sense for users of both the overall contributions of different delegations, and also a sense of how the level of compromise within the process, as delegations saw large numbers of their proposals rejected. A separate display allows the success or failure of different delegations during particular votes, presented as a spine chart, allowing users to gain a sense of whether delegations tended to be on the winning or losing side of controversial measures. This display
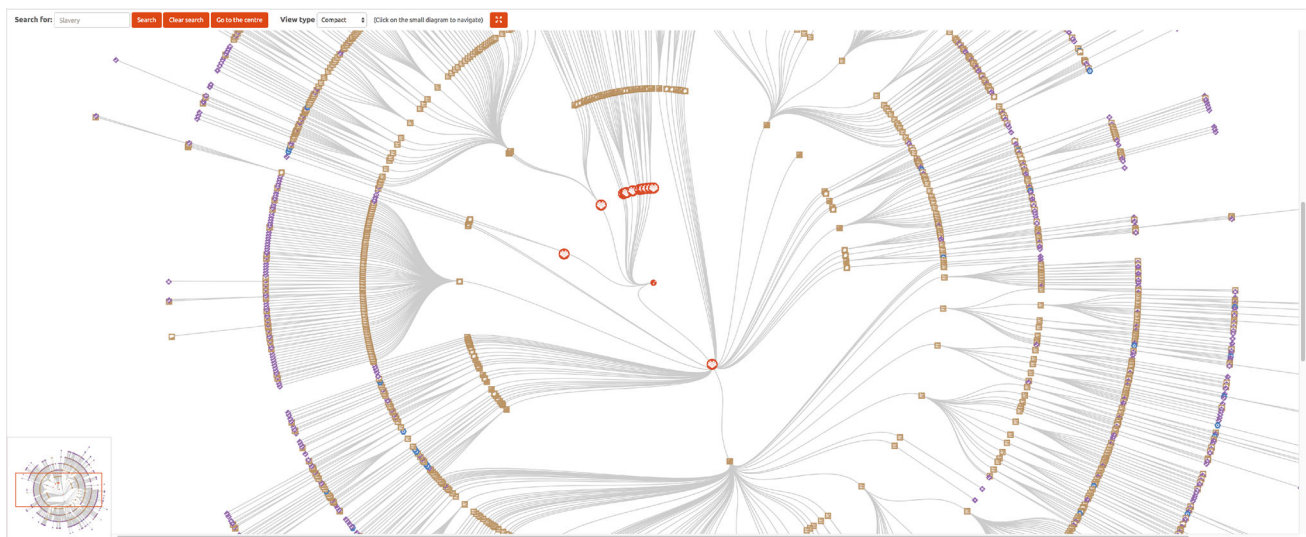
Fig. 7 A summary page illustrating the work of all of the committees of the Convention, in this case showing just the formal relationships between proposals, rather than the sequence of events
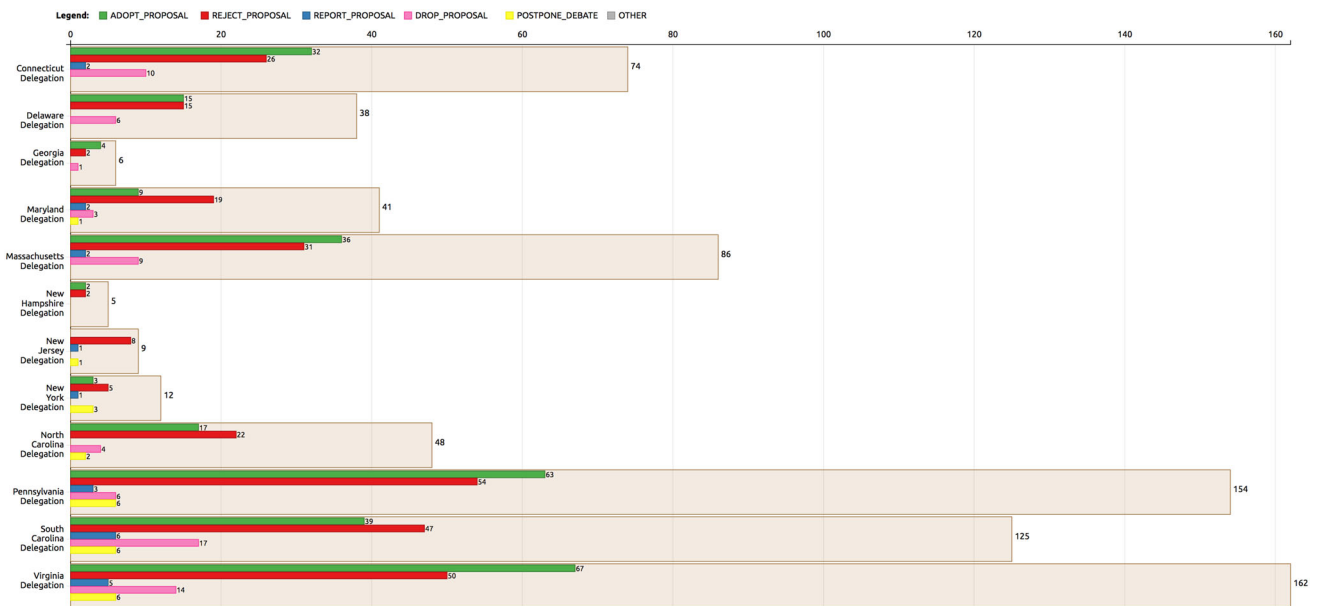

Fig. 8 A display showing the proposals made by different delegations and whether they were accepted or rejected

presented a particular challenge, since for large numbers of votes within the Convention, and especially towards the end of the process, there is uncertainty as to the way particular delegations voted, even if the outcome of the vote is known. The spine chart therefore includes markers to show the level of uncertainty and also a display of any abstentions.

9 Use of this model in practice

The overwhelming majority of effort, as far as data capture was concerned, was to convert the records of the Convention into a timeline of specific events. This work involved a combination of data entry and an interpretation of the records, in two senses. Descriptions of a proposal to be debated needed to be converted into the precise change to the texts intended, and some inconsistency between the extant records (usually to do with the precise sequence of events) needed to be resolved. Those converting the records into this model read through the parallel records of each committee session (where multiple records existed), and decided how to reconcile any conflicts between the records. They recorded such decisions in a private "editors' commentary" as part of the process. It was initially envisaged that reconciling the records

in this way might be impossible—especially if competing forms of words were found recorded for the same proposal in different sources. The platform was therefore designed to allow for competing versions of the Convention timeline to be captured, and for the platform to be able to capture and display any uncertainty about the precise wording of the texts in question. However, it was found that in practice these features were not needed and that (where records existed at all) it was always possible to reconstruct the timeline of particular sessions if records were carefully reconciled by subject experts. An ability to read the records closely and to interpret them in a consistent and logical way was the key to making this part of the project a success. The advantage of designing a platform for ease of use by non-technical users was that that those recruited for data entry could be selected for their subject matter expertise. Due to the nature of the material and the model, an automated process for ensuring the accuracy of data entered into our system was not possible. We implemented a system that would allow the data entry for each session to be marked as verified and for those in charge of the project to view who had entered and who had checked each section of the data.

This data entry phase also involved a certain amount of formative evaluation. As data were entered, we continually assessed whether the model as designed (which had been developed principally from a study of parliamentary manuals) was capable of capturing the work of a process of negotiation from the records that survived in practice. The process of data entry provided multiple opportunities to consider how the model would be best used to most accurately capture the process of decision-making, and the decisions taken about how to use the model were documented to ensure consistency. Where appropriate, minor changes to the data entry dialogues were made to encourage proper use.

Experience proved that data entry was intuitive for non-technical users with a few hours of training and supervised practice. The most frequent type of event in our model of the Constitutional Convention debates is the "Document Amendment". Those working to enter the data select the point in the timeline where they wish to insert an amendment. They then select the document they wish to amend, and whether they are amending the base document or one of the proposed amendments. Once they have made this choice, the platform presents them with the current state of the text at that moment, which they are invited to edit to reflect the state of the document as it would be if the new amendment were to be accepted. They also enter other information, such as the source from which this event is taken, a free-form description of the event, and any known proposers of the amendment. When they have finished, the platform calculates the difference between what the user was presented and what they returned, and associates that patch with the new event.

The next most frequent type of event is a "Decision Event". This records a decision on a particular proposal, be it to alter the text of an amendment, to adopt a section of text into a document, or to accept or reject a document as a whole. There was considerable inconsistency in the records as to the level of detail with which such decisions were recorded. Sometimes the records note with certainty which delegations voted for or against particular motions; sometimes, only the totals on each side were known; sometimes, only the outcome was known. Again, it was feared that the extant records might provide conflicting accounts, and so the platform was designed to allow competing accounts of the votes on particular questions to be displayed, or simply to represent the uncertainty created by the records themselves.

It was sometimes necessary to infer from the records that a particular decision had been taken. For example, it was the practice of the Convention to debate and amend sections of text and then to approve or reject the amended section as a whole. Sometimes, this approval is not recorded in any of the extant sources. This may reflect the fact that the Convention was inconsistent in applying its own procedure, but it is equally likely that a unanimous consent to accept a section as amended and move on to the next order of business was simply not recorded by the secretary as such. Our interface, however, required the insertion of "Decision Events" to capture what the editors inferred to be a decision to agree text and move on. Such interpolated events are clearly marked within our platform. The need to include them highlights the fact that this project produces a model of negotiations, not a literal transcription of source material. It may be obvious from the sources that a particular piece of text has been agreed, even if there is nothing in them explicitly stating the fact.

The implicit rejection of text is a little more complicated. It will be the case in the course of a negotiation that particular suggestions that have been scheduled for debate have simply been overtaken by events—the section of the document to which they refer may have been altered in ways that make the suggestion redundant, or a similar suggestion may have been debated and agreed. In some cases, debate on an issue may simply be managed in such a way that a formal conclusion is never taken, perhaps to avoid the embarrassment of those involved. In these situations, there may never been a formal rejection of a proposal, and even to infer one at a specific point in the timeline may be misleading. For this reason, as well as marking proposals as "accepted" and "rejected", the *Quill* platform's model includes the ability to mark a proposal as "dropped". From the point of view of the model, this has an identical effect to marking a proposal as rejected. It is removed from the list of pending proposals, along with any child amendments, and it is not incorporated into the document. However, including this as a specific type of event allows a more accurate representation of the process of negotiation than the simple binary choice of accepting or

rejecting a proposal, and can be made visually distinct for users.

The most surprising aspect of the platform for new users is that most documents debated by the Convention need to be represented at least twice. Most committees do not work from a blank sheet of paper, but work from an initial base text, either suggested by one of their members or passed to them from another committee. Frequently, they work through this document line by line or paragraph by paragraph, and in so doing produce a new report. The Convention operated in the following way: a framework set of proposals, or suggested document (such as the famous "Virginia Plan") would be offered to the Convention. This would be referred to a subcommittee—in the case of the Virginia Plan, the whole Convention sitting as a subordinate committee. This committee would work through the document section by section and clause by clause, and produce a report for the Convention to consider. The Convention would then work through this report, again amending section by section and clause by clause. In this way, everything would have been considered at least twice, once by each committee.

In a world of paper, quill, and ink, this process created a significant record-keeping challenge, which it would have been the task of the secretary to manage. As the Convention or subcommittees worked through the documents referred to them, he would have had to write out the new text on clean sheets of paper. No doubt these sheets of paper rapidly became untidy and even hard to follow, and perhaps, it is for this reason that they were not entrusted to Washington for safe-keeping but were instead deliberately destroyed, even though copies of the various base documents are extant.

When represented in the *Quill* platform, this process looks identical. If a committee is working through one document to create its own report, the initial document is not shown as amended, but rather the clauses of it are modelled as being gradually incorporated into a new document, which represents the report of the committee. The platform captures the relationship between these documents by allowing any document in the platform to be marked as having one or more "ancestor documents". In visualizations currently under development, this allows readers to view the overall changes made by a committee through this process of revision.

We have found that training is required to help users to interpret the journals of negotiations in order to build successful models within the *Quill* platform. Modern users, even those with experience of research in political and legal history areas, are relatively unfamiliar with the details of parliamentary-style processes. It is important to emphasize to new users that the platform models negotiations, rather than demands a one-to-one relationship between the events we record and wording within the records of a negotiation. We record, for example, the decisions taken upon a proposal,

even where those decisions must be inferred from the records rather than corresponding to a particular vote. In some cases, the best use of the model must show an understanding of the process as it was understood by participants, and requires editors to decide on a consistent way to represent particular situations. For example, where a committee works through a section of a proposed text clause by clause taking a vote on the section as a whole only once that section is complete, this is most accurately modelled as a blank amendment (or one containing only the section number) to represent the moment when the committee turns its attention to a particular section, followed by sub-amendments representing the consideration of each clause. This allows the committee's sense of its own work to be accurately captured in our model, and groups correctly related proposals, but it requires those entering the data to be trained in its use and to apply it consistently. It is not possible (or desirable) to enforce such use in software. Another example is where a clause is "subdivided" for separate votes, as many parliamentary-style processes demand. Our data entry manual recommends that this be modelled as a proposal to reject the original wording and then the introduction of two new amendments representing each part of the divided clause. Our experience has been that where these modelling decisions are correctly explained to users during training, they rapidly become natural to those undertaking data entry tasks.

We were also concerned to make it easy to encourage precise commentary and accurate record-keeping. For users with appropriate permissions, a button to add commentary to any event within a timeline was presented, which presents the user with a pop-up form. Editors creating commentary collections use this button to add their comments. The data entry team used this same system to flag issues within the timeline which required review. Commentary collections serve two functions. The first is to display commentary relevant to particular events or other objects within the platform. The second is to provide a mechanism to guide users to points of interest within the timeline. Collections dealing with particular topics serve as both guides and bookmarks for readers, as well as providing explanatory notes. Since several authors may choose to comment on the same objects for different purposes or offering distinct interpretations, this also allows the platform to present multi-authored explanatory material while making clear the purpose and authorship of particular comments.

As we have worked with wider ranges of users and on different material, we discovered that the original data entry screen proved unnecessarily detailed for the vast majority of uses, and focused attention away from the texts under discussion at a particular moment. The ability to create events at arbitrary positions within a timeline is powerful, but frequently unnecessary. Teams of users entering data into the platform spend most of their time adding events sequen-

tially as they work through particular journals of negotiations. These users spend most of their time adding events to the end of the timeline. For this reason, we have created two screens to assist them. The first of these displays, which we call the "document library", shows the documents available to a committee at the end of its timeline, divided into documents that have been agreed, documents that are still being debated, and documents that have been referred to other committees for review. For each of these documents, a link to a single-page display of the current agreed text of the document is available, as is a link to the place in the timeline where the document was created and to the place in the timeline where the last decision about the status of the document was made. This allows users to quickly see the current state of negotiations. A second display, which we call the "Committee Secretary's view", again focuses on the state of the documents available to the committee at the end of the timeline (see Fig. 9). This screen combines elements of the original editing screen and the session visualization screens, allowing easy access to the list of documents and proposals currently before the committee, and easy access to the accepted, intermediate, and proposed texts of any proposal, including a view highlighting the specific changes that it would make. From this screen, it is possible to add events only to the end of the current timeline, rather than to make more arbitrary changes. This simplification allows the available space on the screen to be devoted to the most frequent tasks of data entry, while access to the more powerful tools remains available. This screen is also more useful when using the platform for record-keeping during live meetings.

A significant problem that we have encountered as we have expanded our work is the nature of the source material available to model various negotiations. In addition to the model of the Convention that is available online, we have a number of other projects in various states of development. We have observed that, although the process of debate within formal negotiations and legislative assemblies is frequently well documented and often formally published, the initial draft texts introduced for discussion are missing from the journals and formal publications, and require additional archival work to recover. The work of subcommittees is frequently not published at all, or published separately from the main journals. Counterintuitively, this suggests that the publication of the proceedings of legislative assemblies and other negotiations have been intended to record the fact that debate had taken place, and memorialize the fact that particular individuals had taken part, but provide in themselves even a careful and diligent reader with insufficient information to recover the details of the negotiations. If initial draft texts ever proved impossible to recover, our model could be adapted to work backwards from final texts rather than forwards from initial drafts, at the cost of significantly complicating both the user interface for data entry and the level of skill required to work

with the records. In practice, having examined the records available for modelling the work of French assemblies after the French Revolution, American state-level constitutional conventions in the nineteenth century, the work of the Paris Peace Conference in 1919, the work of the United Nations, and the framing of India's constitution, we assess that while the necessary records frequently need to be collated for the first time from a variety of sources, they are nevertheless likely to be available in most cases in sufficient detail and with sufficient providence to allow our model to work.

10 User evaluation

We built the core of the platform and web interface over the course of the academic year 2015–6. The platform was then opened to three interns employed at Utah Valley University's Constitutional Studies Center. These interns were given several days of training via video link and then encouraged to experiment with the platform. They were encouraged to try modelling parts of the 1787 Convention records using the platform. They were able to accurately and consistently use the model, and highlighted a number of deficiencies in the user interface for both data capture and readers. A detailed record of their observations and suggestions was kept and used to inform modifications to the platform and a set of editorial conventions that would be used to model the records of negotiations consistently. Although other team management tools were initially used, it eventually became apparent that the *Quill* software itself provided by far the best record of this set of editorial decisions.

From June to October 2016, a recent Oxford University graduate in history (a co-author of this paper) was employed to do the work of the data entry for the Convention, using the 1911 publication of the records, and a graduate in law was employed at Oxford from September to December to assist with verification. We were surprised by the complexity even of the short negotiation covered by our case study. We had originally guessed that this would require around five hundred events, but the final model of the Convention required close to four thousand. Interns at Utah Valley University continued to assist with verification, with some of the ancillary and less historically difficult data entry tasks. Data capture for the Convention was completed in approximately three and a half months in total. We estimated that approximately a day in total data capture was required for each day of the Convention, though this time varied significantly depending upon the state of the records. None of those working on the data entry or verification required specific technical training beyond a few days' training on the specifics of the *Quill* platform and a discussion of the assumptions of the model. Throughout the process of data capture, a record was kept of any aspect of the Convention records that had been diffi-

Fig. 9 "Committee Secretary's view", allowing data entry for the majority of tasks and a less cluttered display

cult to model and the ways in which those difficulties were resolved. At the end of this process, the two graduate students involved in the process of data capture wrote a detailed data entry guide, which incorporated the decisions taken by the editorial team during the process as particular circumstances had been encountered. This data entry guide is intended for use in future work.

We subsequently held a workshop to evaluate both the reader interface and the interface for data capture. In the workshop, we invited six users unfamiliar with the platform to explore the records of the Convention and to attempt to encode one day of negotiation which was part of the process for creating the Universal Declaration of Human Rights [44]. The (unpaid) participants in this workshop were unfamiliar with the platform, came from a range of disciplines, ranged in education level from graduate students through to senior researchers, and were recruited for their interest in digital humanities or interdisciplinary projects. After an hour of introduction to the platform, all of them were able to grasp the basic use of the platform from the reader's perspective and understand the basic conventions and metaphors used by the web interface. All of them were able to understand how to create the basic records related to the work of a process of negotiation. Most of them were able to translate the documents given into an accurate model of the start of the day of negotiation given. In feedback, users commented that the hardest part of this process was understanding the conventions of the minutes in question, not the use of the technology they had been given. This confirmed our view that, as far as data entry is concerned, our platform was appropriate for use by domain experts working within properly agreed guidelines. Our evaluation of this workshop was a mixture of observation of user behaviour (both in the workshop itself and in the subsequent analysis of user behaviours on the website using analytics tools) and evaluation of a questionnaire. Users were asked to undertake both specific and open-ended, self-guided tasks, and to do so with minimal assistance once an introduction to the platform had been given.

We have invited several domain experts to prepare commentary collections for us on a volunteer basis relating to specific questions raised by the records of the Convention. We have already published on the web the first of these, by Lindsay Chervinsky, who was completing her doctorate on the idea of a President's cabinet.[1] Those who have no other experience of data entry have found it easy to attach commentary to specific objects within our timeline. The visualizations within the platform, and the reconstruction of the paperwork available to members of the Convention during particular moments of discussion, have helped them to offer clearer explanations and more succinct descriptions of the progress of particular debates and the reasons for particular outcomes.

During the summer of 2017, we held a week-long workshop for five invited representatives from the Library of Congress, the Bill of Rights Institute, and other non-profit organizations with experience of assisting American high school teachers to develop digital resources suitable for use within their classrooms and conforming to the needs of various state education standards, funding travel, and accommodation expenses. These users were selected for their familiarity with the records, with the requirements of curricula in the USA and for their experience assisting educators with the preparation of materials for classroom use. We monitored both guided and unguided uses of our public website, and gathered written feedback from the participants. All of the participants had some knowledge of the work of the Convention, but no prior experience with our platform. As in previous workshops, this workshop was organized around both specific and open-ended tasks. Feedback was obtained through observation of user behaviour, a questionnaire, and by asking participants to write their own evaluation of the platform. This feedback was extremely positive, both in the sense that users reported enjoying using the platform (as might well be expected) and in the sense that users were both able to complete the tasks specified and to articulate ways that they would be able to integrate the platform into teaching, but confirmed our view that simplified and guided interfaces will need to be developed to make the platform useful for classroom use, and we will be implementing a number of their suggestions over the coming year. All of their feedback emphasized that the platform enabled the richness of the existing records to be better understood. Aside from the reconstruction of the texts themselves, the visualizations of the process and other analytical tools offered by the platform enabled the users to rethink the importance of formal process in supporting the successful conclusion of such negotiations. The model offered of the 1787 records poses a particular challenge to one way in which the work of the Convention is currently taught, which is to encourage students to see the outcome of the Convention as essentially a compromise between the rival plans offered by the delegations from Virginia and New Jersey, both offered early in the Convention. Though the final document produced by the Convention mixed ideas present in both plans, the process by which this compromise was reached involved detailed discussions of the one over several iterations, and relatively little discussion of the other. All participants in this workshop commented that the design of the platform in its current iteration enabled them to focus on the substance of debate and to better understand the significance of process, enabling them to see the final constitution as the result of many actors making a variety of contributions rather than a smaller number of significant personalities.

One particular issue that has become clear as a result of these evaluations is a lack of familiarity with the principles

---

[1] https://www.quillproject.net/commentary_collections/7.

of parliamentary process which greatly hinders the use of the platform. New users effectively have to learn both to use an unfamiliar platform and at the same time to master the basics of parliamentary procedure in order to understand the information they are being given. We suspect that this is a difficulty common to many projects that have used official records to visualize the behaviour of legislative groups. We are developing a variety of materials (including interactive tools) to explain parliamentary process. After experimenting in workshops, we have found that even a 10-min exercise in a group working to amend a short piece of text greatly clarifies the information that the platform captures and displays. Based on this, we are working on a mixture of explanatory text, diagrams, videos, and interactive exercises that can guide users to a better understanding of the basic principles of formal negotiation. Nevertheless, all aspects of the platform currently require too much effort on the part of new users. Our evaluations show that while a motivated user rapidly learns to extract detailed information from the visualizations we present, and can rapidly develop a much more nuanced understanding of formal processes of debate, the combination of unfamiliarity with the processes being visualized and the need to learn to read new visualizations means that it takes some time to make good use of the platform. The users that have made the most rapid progress are the ones that had a pre-existing familiarity with the records being presented, or at least with parliamentary-style processes. Teams undertaking data entry tasks require significant training to understand how to interpret parliamentary records, but once this training is given, rapidly learn to use the data entry tools rapidly and consistently.

10.1 Other feedback and suggestions

One challenge posed by some of our target users was how to integrate digital material with existing classroom material, and in particular with lesson handouts that might be distributed on paper. We judged that URLs would prove too cumbersome for classroom use given these constraints. We therefore designed a "quick jump" system that would be intuitive for both handout creators and pupils. Most material in the platform displays a code consisting of a letter and a number. Entering this code in the box on the title bar of any page will take users directly to the screen required, whether a commentary collection, the visualization of a particular session, or any of the other views offered by the platform. On quick jump-enabled pages, handout authors can obtain either an image containing this code or HTML code suitable for embedding in other web pages. This has proved to be an effective solution to this problem and has become the preferred way for users to direct one another to particular parts of the platform.

We have been asked by various researchers to make adjustments to the platform to allow it to be used for a wider range of material (especially that relating to foreign language material) and to be used to capture debates in real time as well as working from historical materials. In the latter case, we believe this can be achieved mostly through hiding options, and in particular the ability to edit the timeline arbitrarily, and that the tools that we have already developed to simplify data entry may meet most of this requirement. We will be conducting workshops to evaluate what further work is needed in this area.

Adopting a suggestion frequently made, we will be extending the platform to improve the machine-readable interfaces. In particular, we will implement an XML output, using TEI conventions [5]. Although this is likely to result in an extremely complicated set of XML documents, it would be suitable for archiving and importing purposes.

The most frequent question asked is whether the model we have developed would need adapting for specific material. As the range of materials expanded, some improvements have been incorporated. The original 1787 material, for example, featured no roll-call, and the original model had no way to record who was present at a particular session separately from recording membership of a particular committee. However, we are confident that the model (based as it was on a study of parliamentary manuals) can be used for most processes of negotiation, with appropriate training for those performing data entry. The model does not itself work to enforce any particular set of standing orders. It therefore does not need to be modified to reflect the different practices of particular parliaments, for example.

11 Conclusions and future work

We have developed a system that allowed domain experts with minimal technical training to model the almost four thousand proposals and votes that resulted in the text of the US Constitution, even when working from imperfect records. The history of the final text of the Constitution can be accounted for within our model, from beginning to end, and although the platform has the ability to present places where the text is uncertain because of conflicts in the manuscripts, there were no discrepancies substantive enough to warrant it for the 2016 presentation of the records. Such variations as there are relate to extremely minor points of capitalization and spelling. The nature of the records was such that they required significant expertise to accurately model. The reconstruction offered would not have been possible from any one of the surviving sources, but taken together and used systematically and rigorously, we believe they capture the complete work of the Convention's formal business for all of the committees where James Madison and William Jack-

son were present. For other smaller committees, we have been able to show the specific text that was given to them to consider and the report that they returned. These areas of darkness, however, are much smaller than might be assumed, and the process of producing this edition has generally reassured, rather than challenged, our confidence in the accuracy and coverage of the extant records.

Our model is not tied to the presentation of these particular materials, but rather intended to be deployed for a wide range of records related to formal negotiations. The model presented here is generalized and appropriate to capture, analyse, and present the records of formal negotiations, where those records record specific suggestions and decisions. We deliberately do not encode in the model the specific rules of debate (such as how many times a person may speak on a question) that would make the model specific to one particular process, and which would, in any case, be enforced by the participants themselves. Rather, the model captures those elements common to parliamentary process over the previous centuries in a way intended to encourage comparative analysis. We are currently identifying targets for future work and will be expanding the platform to assist with the presentation of multi-language texts. Expanding the range of materials held within our database will enable us to both quantitatively and qualitatively compare and contrast different processes of negotiation. As the library of negotiations modelled in our platform expands, we are developing tools to compare processes of negotiation, comparing their complexity and character. If they are to be meaningful, such comparisons require that our model be used consistently. We will continue to refine both the explicit training available and the behaviours that our user interface encourages.

We are also keen to make the material we are presenting useful in a classroom setting, and especially in the classrooms of the USA. We are currently in discussions with non-profit organizations that work to generate classroom material to highlight several potential approaches. Firstly, we would need to provide an easy way for content creators to integrate our material into their existing lesson plans. Secondly, we would need to provide them with an interface that would let them create resources suitable for classroom use within our platform. A particular challenge is that in many classroom settings, students are using tablets or even smaller mobile devices, and the current interface was not designed for very small displays. We are aware that new (and greatly simplified) interfaces will be necessary for smaller screens, perhaps ones that emphasize summary data rather than the ability to navigate the level of detail offered by the full interface.

Early in the development of the platform, we built in the capability to attach keywords to individual debates. Deciding on the correct keywords and attaching them to events are currently one of the most labour-intensive parts of the data entry, and the least useful feature of the platform for its most expert users. However, our workshops have shown that less expert users and new users value the keywording system as a way to discover data within the platform. We are therefore investigating whether any natural-language processing techniques could enable us to automate this keywording process, noting that many events involve changes to very small parts of texts (that is, changes in only a few words) and that any automated process may in any case not capture the political significance or implications of a small change. We are hoping to apply lessons from the semantic tagging SAMUELS project [35] and the work of the Hansard Corpus project [14] which applied these techniques to British parliamentary records to assist with developing automated keywording mechanisms suitable for *Quill* project libraries.

As noted above, our current implementation manipulates the negotiated documents themselves as plain text due to limitations inherent in the diff–match–patch algorithm that make it impossible to use it reliably with any kind of markup language. However, we are working on a replacement system for text processing that will replace this with an algorithm capable of storing structured documents. This will enable the platform to offer better representations of more modern documents and to integrate the markup standards adopted by other projects, at the cost of requiring a more complicated interface for data entry and those entering data to become familiar with the details of such markup.

We believe that this platform is relevant to the presentation of records held on the negotiation of treaties, constitutions, and innumerable pieces of legislation created in parliamentary settings. We believe that this platform has the potential to democratize understanding of these complicated processes and transform the utility of existing digitized collections for a wide range of audiences.

References

1. Beelen, K., Blaney, J., Blaxill, L., Cochrane, C., Hirst, G., Kamps, J., Marx, M., Naderi, N., Seaward, P., Steer, M., Winters, J.: Digging into linked parliamentary data (DiLiPaD) (2016). http://repository.jisc.ac.uk/6544/1/DiLiPaD_final_report_1.pdf. Accessed 28 Apr 2018

2. Bilder, M.S.: How bad were the official records of the federal convention? George Wash. Law Rev. **80**(6), 1620–1682 (2012)

3. Bilder, M.S.: Madison's Hand: Revising the Constitutional Convention. Harvard University Press, Cambridge (2015)

4. Bundes-Git: Bundes-Git (2018). https://github.com/bundestag/gesetze. Accessed 28 Apr 2018

5. Burnard, L., Jannidis, F., Pierazzo, E., Rehbein, M.: An encoding model for genetic editions (2010). http://www.tei-c.org/Activities/Council/Working/tcw19.html. Accessed 27 Jan 2017

6. Center for American Politics and Public Policy, University of Washington: Legislative explorer: data driven discovery (2018). http://legex.org/. Accessed 28 Apr 2018

7. Digital Variants: Digital variants (1996). http://www.digitalvariants.org. Accessed 4 Feb 2017

8. EDSITEment: The constitutional convention of 1787 (2016). https://edsitement.neh.gov/curriculum-unit/constitutional-convention-1787. Accessed 29 Jan 2017

9. Elkins, Z., Ginsburg, T.: Comparative constitutions project (2005). http://comparativeconstitutionsproject.org/. Accessed 29 Jan 2017

10. Farrand, M.: The records of the Federal Convention of 1787 (1787). http://oll.libertyfund.org/titles/1785. Ed. Max Farrand (New Haven: Yale University Press, 1911). 3 vols. Accessed 30 Jan 2017

11. Farrand, M.: The Records of the Federal Convention of 1787. Yale University Press, New Haven (1911)

12. Fraser, N.: google-diff-match-patch: Diff, Match and patch libraries for plain text (2006). https://code.google.com/p/google-diff-match-patch/. Accessed 27 Jan 2017

13. GNU RCS: GNU RCS (2013). http://www.gnu.org/software/rcs/rcs.html. Accessed 27 Jan 2017

14. Hansard Corpus: Hansard Corpus: British Parliament, 1803–2005 (2016). https://www.hansard-corpus.org/. Accessed 15 Sept 2017

15. Hatsell, J.: Precedents of proceedings in the House of Commons: with observations. In: Precedents of proceedings in the House of Commons: With observations. L. Hansard and Sons, London (1818)

16. Jefferson, T.: A Manual of Parliamentary Practice. Davis & Force, Washington (1820)

17. Kirschner, P.A., Buckingham Shum, S.J., Carr, C.S. (eds.): Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer, London (2003)

18. La Fabrique de la Loi: La Fabrique de la Loi (2018). https://www.lafabriquedelaloi.fr/. Accessed 28 Apr 2018

19. Liberty Fund: Liberty Fund (2016). http://www.libertyfund.org/. Accessed 30 Jan 2017

20. Library of Congress: Part of: James Madison Papers, 1723–1859 (2016). https://www.loc.gov/search/?fa=partof:james+madison+papers,+1723-1859:++subseries+5e,+madison%27s+original+notes+on+debates+in+the+federal+constitutional+convention,+1787. Accessed 29 Jan 2017

21. Lilian Goldman Law Library: Notes on the debates in the Federal Convention (2008). http://avalon.law.yale.edu/subject_menus/debcont.asp. Accessed 29 Jan 2017

22. Linder, D.: The Constitutional Convention 1787 (2016). http://law2.umkc.edu/faculty/projects/Ftrials/conlaw/convention1787.html. Accessed 29 Jan 2017

23. Lloyd, G.: The Constitutional Convention (2016). http://teachingamericanhistory.org/convention/themes/. Accessed 29 Jan 2017

24. May, T.E.: A Treatise upon the Law, Privileges, Proceedings and Usage of Parliament. Charles Knight & Co., London (1844)

25. McNamee, R.: Electronic Enlightenment Project (2016). http://www.e-enlightenment.com/. Accessed 29 Jan 2017

26. Mercurial: Mercurial (2016). https://www.mercurial-scm.org/. Accessed 27 Jan 2017

27. Ministry of Parliamentary Affairs, Government of India: Manual of Parliamentary Procedures in the Government of India (2017). Accessed 4 Feb 2017

28. Myers, E.W.: An O(ND) difference algorithm and its variations. Algorithmica **1**(1), 251–266 (1986)

29. National Archives Catalog: Journal of the Federal Convention (2016). https://catalog.archives.gov/id/7347105. Accessed 29 Jan 2017

30. Parliament: Bills before Parliament 2016–2017 (2016). http://services.parliament.uk/bills/. Accessed 29 Jan 2017

31. Reed, C., Palau, R.M., Rowe, G., Moens, M.F.: Language resources for studying argument (LREC'08) (2008)

32. Reed, C., Wells, S., Snaith, M., Budzynska, K., Lawrence, J.: Using an argument ontology to develop pedagogical tool suites. In: Blackburn, P., van Ditmarsch, H., Manzano, M., Soler-Toscano, F. (eds.) Tools for Teaching Logic. Lecture Notes in Computer Science, pp. 207–214. Springer, Berlin (2011)

33. Riker, W.H.: The heresthetics of constitution-making: the Presidency in 1787, with comments on determinism and rational choice. Am. Polit. Sci. Rev. **78**(1), 1–16 (1984)

34. Robert, H.M.: Parliamentary Law. Century Co., New York (1923)

35. SAMUELS: SAMUELS (Semantic Annotation and Mark-Up for Enhancing Lexical Searches) (2016). http://www.gla.ac.uk/schools/critical/research/fundedresearchprojects/samuels/. Accessed 15 Sept 2017

36. Shillingsburg, P.L.: Scholarly Editing in the Computer Age: Theory and Practice, 3rd edn. Editorial Theory and Literary Criticism. University of Michigan Press, Ann Arbor (1996)

37. Silverbrook, J., Johnson, M.: ConSource (2007). http://www.consource.org/. Accessed 29 Jan 2017

38. Software Freedom Conservancy: Git (2016). https://git-scm.com/. Accessed 27 Jan 2017

39. SRI International: Structured evidential argumentation system (SEAS) (2018). https://www.sri.com/engage/products-solutions/structured-evidential-argumentation-system-seas. Accessed 28 Apr 2018

40. Surden, H.: Constitution Explorer (Beta) (2015). http://www.harrysurden.com/projects/visual/USCode_D3/constitution/US_Constitution_Tree.html. Accessed 29 Jan 2017

41. Syrett, H.C. (ed.): The Papers of Alexander Hamilton, vol. IV. Digital edn. University of Virginia Press, Rotunda (2011)

42. Tauberer, J.: Bills and resolutions (2004). https://www.govtrack.us/congress/bills/. Accessed 29 Jan 2017

43. The Library of Congress: United States Constitution (2016). http://www.loc.gov/rr/program/bib/ourdocs/Constitution.html. Accessed 4 Feb 2017

44. United Nations: Universal declaration of human rights (1948). http://www.un.org/en/universal-declaration-human-rights/. Accessed 27 Jan 2017

45. U.S. Government Publishing Office: House rules and manual (2016). https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=HMAN. Accessed 27 Jan 2017

46. Vile, J.R.: The critical role of committees at the U.S. Constitutional Convention of 1787. Am. J. Legal Hist. **48**(2), 147–176 (2006)

47. Walton, D.: Visualization tools, argumentation schemes and expert opinion evidence in law. Law Probab. Risk **6**(1–4), 119–140 (2007)
48. Winograd, T., Flores, F.: Understanding Computers and Cognition: A New Foundation for Design. Addison-Wesley Longman Publishing Co., Inc., Boston (1987)