Moving Early Modern Theatre Online: The Records of Early English Drama introduces the Early Modern London Theatres Website

By Tanya Hagen, Sally-Beth MacLean and Michele Pasin
Records of Early English Drama, University of Toronto/Department of Digital Humanities, King's College London

1.The Context

The Records of Early English Drama project is an interdisciplinary research and editorial project based at the University of Toronto. REED was founded in 1976, its primary purpose being to find, transcribe and edit for publication surviving records of drama, music and popular mimetic entertainment before 1642, when the Puritans closed the public theatres in London. Thanks to the efforts of a dedicated staff and determined editors in Canada, the US and UK, the project is still going after all these years, a hardy veteran of collaborative humanities scholarship. The list of print publications now totals twenty-seven collections in thirty-three volumes, with a landmark collection for the *Inns of Court* published in 2011, the second of several for the historic city of London and its neighbouring counties (see Map of REED Collections).

Our first steps to move REED online were taken just over ten years ago, when dedicated funding made possible the development of our first research and educational web site, *REED Patrons and Performances* (http://link.library.utoronto.ca/reed/). The site results from a long-standing wish, on the part of early theatre historians, to trace the activities of professional performers of all kinds who toured to the towns, monasteries and private residences of provincial England. Indeed, this wish was a major motive behind the founding of REED and the extension of its time frame beyond the suppression of biblical cycle drama in the 1570s into the early seventeenth century.

Nine years into the life of the project, Toronto staff began, systematically, to assemble information about the patrons under whose names many of these performers

travelled. Some of the questions that drove our research were: 'Who were they? When and where were they born, and where did they live? Who were their families, titles, connections, spheres of influence?' Where did their entertainers travel and perform? How much did the performers earn and whom did they please or offend? Out of the need to organize and maintain this host of details, the REED Patrons database was born, primitively in Basic, then migrated into dBASE II and IV, and eventually into ACCESS. These programs were never user-friendly and although we made diplomatic noises about opening the resource to other scholars if they visited the office, little use was actually made of the data in this form.

In 1998 Sally-Beth MacLean and Alan Somerset envisioned a more accessible future for this REED data, as an adjunct to what most members of REED's Executive Board viewed as the project's core activities. So long as we did not encroach upon the actual—or even potential— funding sources for production of the print volumes, we could engage in our digital play. Grants from the University of Toronto, the University of Western Ontario and the Social Sciences and Humanities Research Council of Canada made possible the research and educational site that has been freely available on the web since 2003. Another significant gift-in-kind came with the offer of help from our now longstanding partner, Sian Meikle, digital services librarian in the Information Technology wing of the University of Toronto Library. Her technical skills contributed to the transformation of our database entry process, expansion of our research possibilities, and migration of the data onto the web in the standard open-source relational database management platform, MySQL.

Another key collaborator has been Byron Moldofsky, chief cartographer in the Department of Geography at the University of Toronto. The Geographic Information System (GIS) map of England and Wales on the web site resulting from this partnership demonstrates at a glance the many medieval and renaissance performance locations identified by REED researchers, together with major routes they might have travelled and the rivers and other topographical features that could have influenced their choice of

itinerary.[1] Linked with the patrons, troupes and performance events databases, it serves as a visual springboard for exploring performance venues and other locations associated with patrons' office titles.

A development team in Toronto led by Jason Boyd has uploaded all the data from volumes published before 2005, the year when patrons and performances data began to be made available on the web site simultaneously with the publication of each volume. Travel grants have also enabled an entirely new database packed with fresh architectural research and images of hitherto unacknowledged performance venues across the kingdom – what we call the alternative theatres of the provinces.[2] Storage is in the hands of the University of Toronto Library, where the databases, middleware, and web site are maintained on a Library production server. REED has been assured of the Library's long-term commitment to maintain, back up, and regularly upgrade software and hardware as appropriate, another crucial contribution to our digital projects.

We are currently moving in deliberate steps towards revolutionizing our production and publication processes in order to deliver forthcoming collections as fully searchable, hypertextual editions. REED, having been a pioneer of complex publication in print, now aims to be a pioneer of complex publication on the web. The integration of born digital REED collections with other online REED resources, both current and projected, will create a unique interdisciplinary research and educational resource.

We have a credible calling card on the web with *Patrons and Performances* as described above, but the source line entries on the Events and Venues pages to REED volumes are merely interim, beckoning to future links with REED electronic texts. It has been clear for some time that the long-term future of the series must be on the web as well as in print. The cost of the volumes as set by the University of Toronto Press has proved too high for most individuals and for many libraries.

---

[1] The historical and cartographic sources of evidence researched by Sally-Beth MacLean for the Interactive Map are listed individually in the online Bibliography.

[2] The *Patrons and Performances Web Site* development and design is more fully described in MacLean's recent essay with Alan Somerset (2008).

As we develop research and educational tools to widen our audience and bring our discoveries to the attention of teachers, students, and the general public, we intend to make available the collections themselves as fully searchable digital editions. REED editions have always been interactive, but clumsily so in their print form. Anyone using the historical Records text presently has to remember to look to the bottom of the page for a textual or collation note; flip backward to the relevant source document description, or forward to the translation, endnote, glossary entry or index, sometimes in a separate volume. A digital format offers infinitely better possibilities for hyperlinks within each collection as well as subject searches across the series. Enhanced flexibility for the edited texts could enable, for example, the option of resorting records from their usual chronological organization into individual manuscript order or by institution. Dynamic mapping of the edited data will become feasible, connecting not only with individual county and city maps in each collection but also with the GIS map on the *Patrons and Performances Web Site*. Links with other open access websites can be projected: for example, the separate Names Index for members of the Inns of Court could be linked with the online *Inner Temple Admissions Database* (http://www.innertemple.org.uk/archive/itad/index.asp).

Several years ago we welcomed a fleeting opportunity to have the first twenty-four volumes in the series scanned and uploaded on the Internet Archive web site (http://www.archive.org/). We were only able to take this step because in the mid-‘90s we negotiated with our publisher to retain our electronic rights to the volumes. At this time the pdf versions of volumes from *York* (1979) through to *Cheshire including Chester* (2007) can be viewed online, but with only limited search capability.

One of our immediate goals is to move toward print and web publication of forthcoming collections in the series, with *Middlesex including Westminster* as our pilot. As always, we are dependent on further funding both to maintain production but also to raise our processes to a new level. The first step has been taken. In 2007 Alan Nelson, editor of *Inns of Court*, was awarded an eighteen-month Digital Humanities Startup Grant from the National Endowment for Humanities to work with us in partnership with the

Library and the electronic publications coordinator at the University of Toronto Press. As an outcome, we have pilot Inns of Court text files converted from REED's standard editorial ASCII-coded markup into TEI-conformant XML populating the new SQL database designed to enable single-source editorial content for REED to generate two distinct products: prepress-ready text suitable for print publication and dynamic digital publication.[3] The database is housed and maintained on a library server. Next to come will be a generalized scholarly editing interface to implement the specific editorial and production practices of REED, ensuring continuity for the project and enabling new ways of working with the records. Full launch of the first online publication will require interface development, design, and testing, as well as editorial and scholarly review.

The principal subject of this essay, however, is a second major research and educational database launched in its first phase in February 2011. In 2007, a research and development team based in England received an Arts and Humanities Research Council grant to fund record office research for primary documents relating to the eight Elizabethan and Jacobean theatres north of the Thames, and to create and design an annotated bibliographic database then known as the *London Theatres Bibliography* (*LTB*) for delivery on the web as an open access resource for wider public use.[4] Michele Pasin at the Department of Digital Humanities (formely known as Centre for Computing in the Humanities), King's College, London has brought his expertise to the technical aspects of our international partnership: for example, by migrating bibliographic data from Endnote into a newly designed database and building user-friendly interfaces to support editorial work**.** In Toronto REED's Bibliographer, Tanya Hagen, is leading the bibliographic research team on content development for the open access web resource now titled *Early*

---

[3] The deliverables can be downloaded from the REED project website (http://www.reed.utoronto.ca/downloads.html) and used under the conditions of a MIT license.

[4] The AHRC project is led by Professor John McGavin (University of Southampton), with co-investigator, John Bradley (Centre for Computing in the Humanities, King's College, London), in partnership with Sally-Beth MacLean (REED, Toronto).

Modern London Theatres, <http://emlot.kcl.ac.uk/>. In the following sections we will give a more detailed description of both the intellectual and technical challenges this new digital humanities project has brought about.

2. What is the *Early Modern London Theatres*?

Early Modern London Theatres aspires to provide its users with a major encyclopedic resource on the early London stage, as well as a comprehensive historiographical survey of the field. In compiling EMLoT, we aim to identify, record and assess transcriptions from primary-source materials relating to the early London stage, as found in secondary-source print and manuscript documents. Our main criterion in distinguishing between a primary- and secondary-source document is chronological: EMLoT's purview stops with the REED volumes (and the closing of the theatres) at 1642. Under this rubric, a primary source is a document produced before 1642, and a secondary source is one produced after 1642. There are, of course, some exceptions here. We make allowances for works known to have existed in some form before 1642, but for which the earliest surviving witness is a post-1642 document. This applies primarily to play texts: many of Thomas Middleton's and James Shirley's works, for example, did not see publication for the first time until the 1650s. There are also a few instances in which later manuscript sources provide us with valuable contemporary evidence concerning the pre-1642 stage. A petition by Elizabeth Heton, William Wintersall, and Mary Young to the Earl of Dorset, filed *c* 1657-8, speaks of a lease entered into some thirty years ago with the Earl's father for an old barn standing in Salisbury Court (Wickham, Ingram, and Berry 2007, 654). In such an instance, where the substance of the record clearly relates to an event that took place before 1642 (e.g., the construction of the Salisbury Court theatre) and provides evidence of major import to the history of the early London stage, we have chosen to relax our chronological parameters.

Within the discrete groupings of pre-1642 primary and post-1642 secondary-source documents, we maintain the broadest possible selection criteria. Any document, in manuscript or print, may qualify as a primary source, as long as it contains matter relating in some way to early London's theatrical scene: court book, parish register, miscellany, religious polemic, broadside, jestbook, play text, title-page. We also maintain a generous interpretation of relevant content. A document need not refer directly to a performance, venue, or person associated with the professional theatre to qualify for interest. Biographical records of family members of known theatre professionals found in parish registers, for example, may prove useful in building a demographic profile of London's entertainment community. Land surveys and court records can supply valuable evidence regarding the sites on which theatres were constructed.

Similarly, any document – manuscript or print, scholarly or popular – may qualify as a secondary source, as long as it supplies a fresh transcription from a relevant primary source. We are interested only in direct transcriptions: there must be evidence that the editor or author of the secondary source is a witness to the original document. We do not collect allusions, paraphrases, or quotations from earlier edited sources. Distinctions here may not always be obvious: a fresh transcription may have been modernized, while many sources will reprint documents in old-spelling from earlier editions. Recent scholarly works, accompanied by a standard apparatus of notes, present the fewest problems here. Earlier and popular works can be more challenging: editors and authors working before the professionalization – or beyond the parameters – of the academy, are less consistent or reliable in identifying the source of transcribed material. Printed primary sources tend to generate the most problems, as editors and authors generally are less rigorous in handling such material: it is frequently unclear whether a transcription derives from an original, a facsimile, or an edited copy. Our policy on material of uncertain provenance has been to err on the side of inclusiveness, identify records based on questionable transcriptions as such, and cull only when we have established a solid case for exclusion.

Coupled with a broad mandate on source material, our interest in ephemeral and obscure authorities will, we believe, distinguish EMLoT as a particularly original

resource. Two classes of material are noteworthy in this respect: (1) late seventeenth- and eighteenth-century published works either largely unknown outside the realm of eighteenth-century literary and historical studies, or no longer consulted as authorities, and (2) the unpublished papers and research of noted theatre historians.

Under the first class of material, Sir William Sanderson's *A Compleat History of the Lives and Reigns of Mary Queen of Scotland &c* (1656) provides our earliest record from a non-dramatic secondary source: a transcription of an unidentified document that records the cost of the "Lord's Mask" at the marriage of Princess Elizabeth. James Wright's pro-theatrical polemic, *Historia Histrionica* (1698), cites voluminously from John Stow's *Annals* in support of the ancient and royally sanctioned tradition of theatre in England, while Luigi Riccoboni's almost completely unknown *An Historical and Critical Account of the Theatres in Europe* (1741) transcribes a passage from James I's 1603 license to the King's Men.

In extending the purview of EMLoT to the unpublished papers of noted theatre historians, we seek not only to complement our survey of printed secondary works, but also potentially to uncover new or unique material, either previously unpublished, or taken from lost originals. To date, we have conducted preliminary surveys of the Edmond Malone and Francis Douce collections at the Bodleian Library, Oxford, and the Charles William Wallace collection at the Huntington Library, San Marino, California.[5] This research is to be reviewed and integrated into EMLoT over the course of the next year.

All forms of transcription are therefore worthy of note: not only those faithful in every respect to the original, but also the excerpted, emended and otherwise adulterated. We may thus consider not only the frequency with which a primary-source document has been published, but also its various treatments over time, and at the hands of different editors. Which documents tend to be preserved whole, and which heavily excerpted; which preserved in facsimile, and which modernized? In constructing (what we believe is) an unprecedentedly detailed account of the extant archive, we hope also to stimulate

---

[5] We owe thanks to the Social Sciences and Humanities Research Council of Canada's International Opportunities Fund for the grant enabling us to conduct this research.

further discussion on the meta-history of that archive. What may the handling of primary-source materials at various periods and from different cultures tell us about changing attitudes toward, or investments in, the phenomenon of "Shakespeare's stage"?

Properly speaking, of course, neither primary- nor secondary-source document is useful to us *per se*. The crux of our interest, rather, is in the relationship between a primary (transcribed) and a secondary (transcribing) document. The job of compiling EMLoT is to describe that relationship within the parameters of an established template. In the first stage of EMLoT creation and development, we identified several thousand such unique relationships. Our current task, effectively, is to fill in the blanks.

As EMLoT has moved from a flat-face database to an electronic platform, these blanks have multiplied and grown in complexity. An EMLoT record properly comprises a number of interlinked files. We begin with the 'record' file, which yokes together a primary- and secondary-source document, and provides data concerning their relationship: the location of the transcription within the secondary document; citation data for the primary document (as provided by the secondary source); brief notes on the treatment of the primary by the secondary source. In anticipation of a later phase in the process, we have also delimited fields which allow the compiler to enter corrected citation data for the primary source; to link the record to a published REED transcription, and to furnish a unique EMLoT transcription (where a REED transcription is not available).

The 'record' file links to three further files: two "document" files – one each for, respectively, the primary and secondary sources – and an "events" file. "Document" files are effectively bibliographical templates and are generic in that they do not furnish data necessary to establish a relationship between transcribed and transcribing documents. A single "document" file may in this respect link to several hundred records, depending on the number of transcription relationships in which it is involved. It is in the triangulation of record and document files that EMLoT serves both its primary function as a bibliographic resource, and an ancillary historiographical interest.

EMLoT becomes an encyclopedic research tool in the form of the "events" file, as it is in this context that we record details of a transcription's contents. A summary furnishes a short-title, an "event type" field allows us to build a keyword profile, and an abstract provides a narrative description of the event. Date fields allow us to discriminate between and specify the dates on which an event happened and was recorded, as well as to identify a relevant feast. Association fields note the people, venues and troupes involved in a given event. Each of the three types of association field is served by an underlying cache of "people," "venue" and "troupe" files. As in the case of the "document" files, a single "events" file may link to several records, depending on the number of transcriptions identified from the same document.

EMLoT distinguishes itself from other REED resources insofar as it does not itself provide fresh transcriptions from archival sources; its purview does not extend beyond London; and, because EMLoT deals in material associated with purpose-built spaces and professional theatrics, its chronological focus is primarily sixteenth- and seventeenth-century. Rather than serve as a bibliographical adjunct to REED's projected London volumes, EMLoT is intended to take its place alongside other electronic REED publications as a discrete, but fully interoperable, resource. Above, we touch on the architecture already in place to link EMLoT 'record' files to eREED volumes; at present, EMLoT "people" files also link to the *Patrons and Performances Web Site* through a "patrons" field. Now that the groundwork of EMLoT is complete, we look toward to establishing a more complex system of pathways to connect EMLoT to other REED Online resources. A current priority, for example, is to link EMLoT to *Patrons and Performances* "events" files. As the programmer initially responsible for developing EMLoT's online architecture, Michele Pasin will address below the greater challenge of making REED's electronic resources talk to each other.

3. EMLoT: towards an interoperable humanistic database

At King's College Department of Digital Humanities we have been involved in the construction of digital scholarly resources for a number of years now. But despite the considerable in-house expertise available, the prospect of building a "London Theatres Bibliography" initially caused a mixture of excitement and apprehension. The strong focus on representing the *transmission history* of the relevant literature, coupled with the goal of storing a large number of facts related to the *people*, *events*, and *venues* in Shakespeare's London, seemed quite a challenging project to undertake. Probably even more challenging, when considering the already established *Patrons and Performances Web Site* mentioned above. In fact, although the *Patrons and Performances* has quite a different take on the subject, nonetheless it stands on the horizon of EMLoT development as a reference point that needs to be built upon and linked to. In other words, we soon envisioned a number of usage scenarios requiring a *unique access point* to the two resources, i.e., a way to query the two databases that relies on their various possible existing points of intersection (e.g., patrons or theatres).

In the following paragraphs we discuss the approaches taken in order to address the specific requirements posed by this novel context. We summarize what technologies have been used and what steps have been taken with the aim of creating a freely available web database. In particular, we will focus on the *conceptual* aspects entailed by the construction of a database for EMLoT world, as we believe this is the key aspect to take into consideration in order to move towards a more interoperable *web of data.* As we will see, this approach is inspired by the discipline of *ontological engineering,* a recent research area in computer science that reuses ideas and methods from philosophical ontology with the aim of building more solid computational data models.

3.1 From Endnote to a MySQL: approach, advantages, technologies used

In September 2008 EMLoT database was only starting to take shape. A lot of research material had already been collected by Tanya Hagen in the form of a very large Endnote library, but soon this type of medium presented several limitations, mainly

deriving from the lack of collaborative editing functionalities and the poorly customizable interface. Consequently the first phase of EMLoT involved extracting of such data from Endnote (http://www.endnote.com) and copying it to a MySQL (http://www.mysql.com) database.

In order to understand better why we needed to transform the original Endnote library into a different format, it is useful to take one step back and ask ourselves, why did EMLoT material need a database at all? Generally speaking, the most important advantage of databases is that they allow us to search for information more easily and efficiently. This is possible because in a database the entities comprising a specific domain have been carefully identified and separated. Another advantage of having all of this data stored in a database is that we can visualize it in different ways, in different mediums, and by different people at the same time.

We can define a database as a structured collection of records or data that is stored in a computer system. The structure is achieved by organizing the data according to a database *model*. The model in most common use today is the *relational* model (RM). As we can see in Figure 1, the most important feature of the RM is that tables are connected to each other according to specific relations. These relations usually represent real world relations: for example the “document” table has a relation “authorship” which points at the “person” table. So, in other words, the key feature of a database is that data get “broken down”, so to say, into smaller units. They are grouped according to certain properties we see in them, and we call each one of these groups a ‘table’. A database is therefore essentially a collection of such tables.

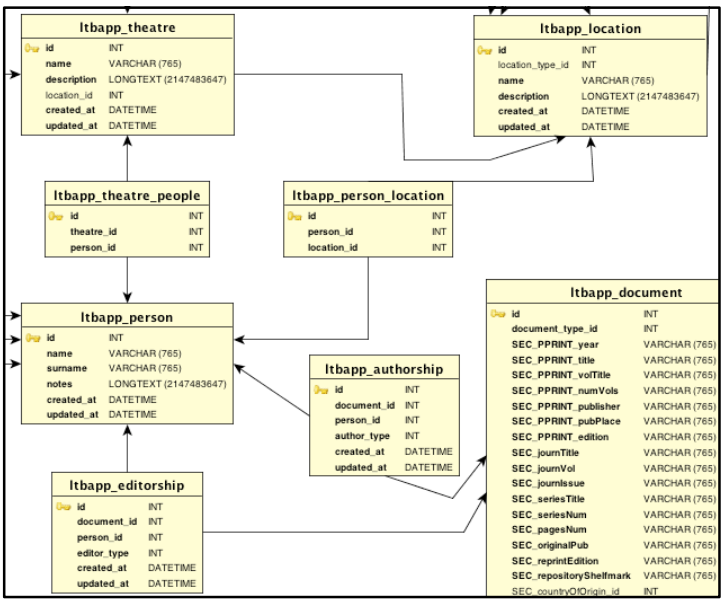

Figure 1: graphical representation of EMLoT relational model

If we take a closer look at the contents of the "theatres" table (Figure 2) we can see that it is composed of a number of records (the rows) and fields (the columns). Each record is representative of an instance of the data-type (entity) we are addressing; instead, a field is representative of one property of that instance. For example, we can see that this table groups data about a theatre's name, a description we can give it, its location, and other meta-information which is not related to the theatre itself but to our action of creating a record.

We can now understand why the data stored in an Endnote library were not sufficient for our purposes: we could say that an Endnote library consists of only one table, containing all the fields related to EMLoT domain. Essentially, what we had to do was to decide how to group these fields in order to create multiple tables (and obviously also a set of relations connecting them all). Now, the interesting question is: how can this be done effectively? Probably one of the most sensible answers is, *in a way that*

id	name	description	location_id	created_at	updated_at
1	Rose		NULL	2009-03-02 1...	2009-03-02 1...
2	Theatre		NULL	2009-03-02 1...	2009-03-02 1...
3	Newington Butts		NULL	2009-03-02 1...	2009-03-02 1...
4	1Fortune		NULL	2009-03-02 1...	2009-03-02 1...
5	Curtain		NULL	2009-03-02 1...	2009-03-02 1...
6	Whitefriars		NULL	2009-03-02 1...	2009-03-02 1...
7	1Globe		NULL	2009-03-02 1...	2009-03-02 1...
8	Phœnix?		NULL	2009-03-02 1...	2009-03-02 1...
9	Court		NULL	2009-03-02 1...	2009-03-02 1...
10	Hampton Court		NULL	2009-03-02 1...	2009-03-02 1...
11	St Peter, Cornhill		NULL	2009-03-02 1...	2009-03-02 1...

resembles the world we are describing. Let us keep in mind that EMLoT deals with records of theatrical events in post-1642 transcriptions of pre-1642 sources. Thus, as previously mentioned, a key requirement was to represent the transmission history of various kinds of documents. Accordingly, *EMLOT* bibliographer Tanya Hagen and database designer Michele Pasin had a number of discussion sessions in which they attempted to make explicit the various "features" of the documents we were going to describe (eg, material properties, publication details, etc). Also, they investigated the extent to which a description of theatrical events was needed in the database, and which are the main "entities" that usually appear in the context of such events (e.g., people, venues or locations).

This process resulted in an initial domain model whose main structure is depicted in Figure 2. Notice how we have three different "poles" that are "orbiting" around the *record* entity at the bottom right of the figure. Essentially, a *record* is an abstract entity representing the process by which EMLoT researcher makes a claim about the *connection* between two documents, ie, when we say that "manuscript document X" has been transcribed in "printed document Y." The logical separation between "record" and '"document" reflects the fact that the same document could have been transcribed many times, in different contexts and with different "styles" (or errors). This conceptual model thus avoids unnecessary duplication of information. Let us now have a look at each one of the "poles" depicted in Figure 2 below:

A) The *source* pole represents all the document-types EMLoT deals with; in general, sources' instances can be transcriptions or originals and are attached to records by means of the *is-about-source* relation.

B) The *person* pole gathers the concepts needed for describing a document's authors and editors, but not only: in fact because we are keeping track of the contents of documents too, very often it is necessary to store information about laymen, players and more generally, people who lived in the period EMLoT is investigating. Thus the person section contains also "historical individual."

C) The *event* pole groups the concepts used for describing happenings of various sorts, but mostly, performance events. These are the events our sources describe. It is remarkable how difficult the 'extrapolation' of event-information from literary texts can be, as the definition and granularity of an "event" cannot be easily agreed upon. In practice, such decisions are made case by case by the editorial team, paying attention to maintaining a consistent approach throughout the entire database (it is worth noting that we have created mechanisms by which an editor can specifically say that some piece of information is the result of his or her interpretation). Finally, the event pole contains also other entities that are less subject to interpretations: *companies* or *troupes* (the "Admiral's Men"), *places* ("Drury Lane") and *venues* ("Clement's Inn").

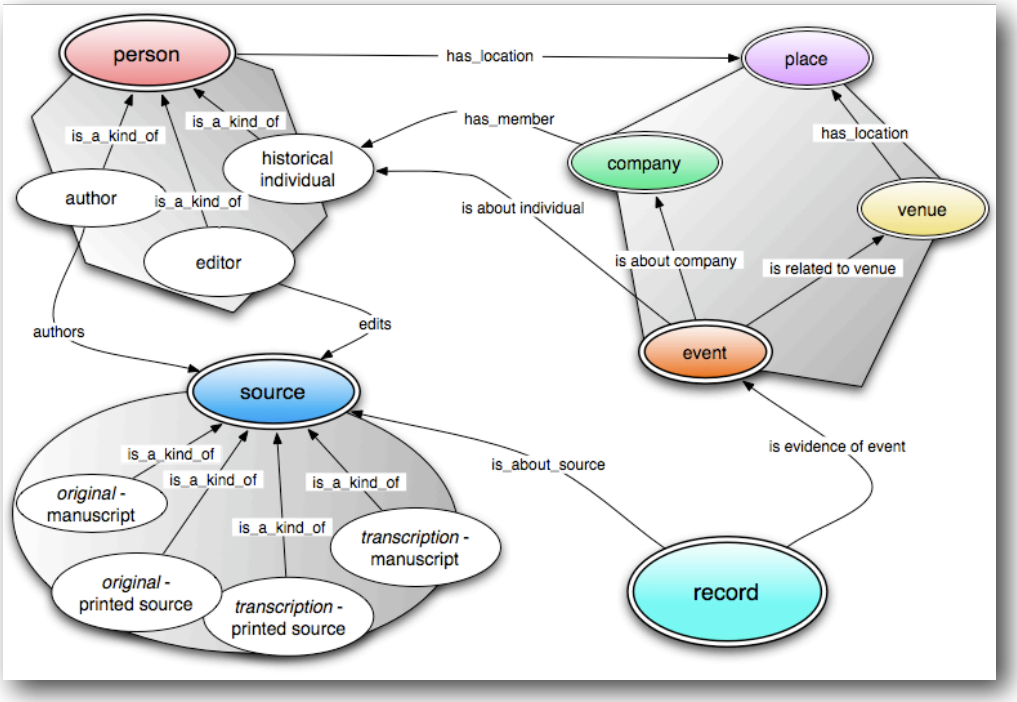

Figure 2. First sketch of a data model for EMLoT application

The conceptual model depicted in Figure 2 obviously represents only the main logical structure of EMLoT world. In order to gain more insight into each of the entities introduced above we performed various other evaluations of the Endnote library data; as a result, we obtained a number of much more fine-grained descriptors such as the ones shown in Figure 3.[6]

[6] An extensive outline of these descriptors will be made available in a separate publication.

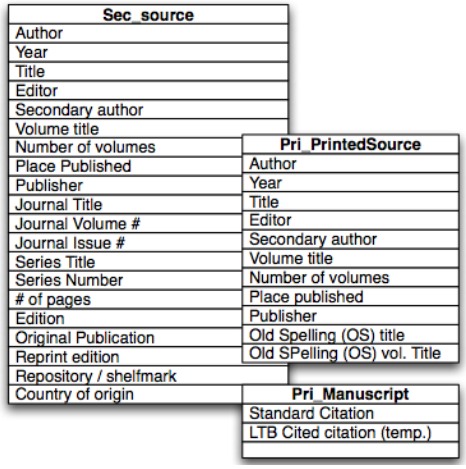

Figure 3. Detailed field descriptors for the 'document' entities

After the conceptual model had reached an adequate degree of stability, we moved on to the implementation work. This involved three different phases:

1) We created a new MySQL database structure based on EMLoT conceptual model.
2) We exported the original Endnote library into the newly created database. Endnote provides a handy "export to XML" functionality, so first we transformed the library to that format for easier processing. Subsequently we wrote a Python (http://www.python.org) script to "explode" that information into our newly created database. This process was not particularly difficult from the purely technical point of view, but we had to work out a number of strategies for spotting identity relations in the XML document. For example, since the Endnote library consisted of only one big table, two documents having the same author exhibit that by having the same person's *name* in the author column. Instead, in the database representation we would have two *references* in the document table that point at the same record in the person table. The main problem here was that Endnote's fields contained various spelling errors or differences that required *ad hoc* algorithms (e.g., for determining that two author' names were effectively referring to the same person, even if they were spelled differently).

3) Once the database contents were in place, we created a web application that allows its visualization and editing. To this aim, we used the freely available Django (http://www.djangoproject.com) framework, a python-based environment that aims at speeding up the development of websites by providing a number of reusable application components. The web application (figure 4) let EMLoT team check the data imported from Endnote, refine it and start adding new ones. After an initial period of testing and familiarizing with the new environment, Tanya Hagen and the other editors stopped using Endnote and continued working with this new system. At the time of writing, the administrative side of the web application consists of more than twenty views addressing the management of the different aspects of the database.

It is important to underline that the process of creating a web database is normally a result of a series of iterations in which new features are added and others are removed, both at the database level and at the interface one. EMLoT was no exception in this respect, so the outline above must be understood as a simplified version of what happened in reality (in particular, see the next section for an example of the refinements we carried out on the initial conceptual model).

Finally, we should mention that what is discussed in this article reflects mainly the first phase of the project, that is, the one involving data capture and representation; a second and equally fundamental phase had to do with the design and construction of adequate presentation mechanisms for the data we collected. A thorough description of all of these aspects would have probably exceeded the scope of the present discussion, and most certainly the space available on this publication, so we decided not to include it here. Let us use briefly mention though that the EMLoT website features both traditional keyword-based search mechanisms and more advanced browsing tools; in particular, a purposely created ‘faceted search’ component (Tvarožek and Bieliková. 2007) allows users to explore the database contents using a highly interactive user interface. To the purpose of gaining more empirical evidence on the effectiveness of these search tools, we are currently running a user-evaluation study of EMLoT at London’s King’s College. The results of this experiment, together with a detailed description of the work done on

the front-end interface, will be made available later this year in a separate publication. For the moment, it is possible to see all of the front-end functionalities in action on the EMLoT website <http://www.emlot.kcl.ac.uk>, which was launched in February 2011.

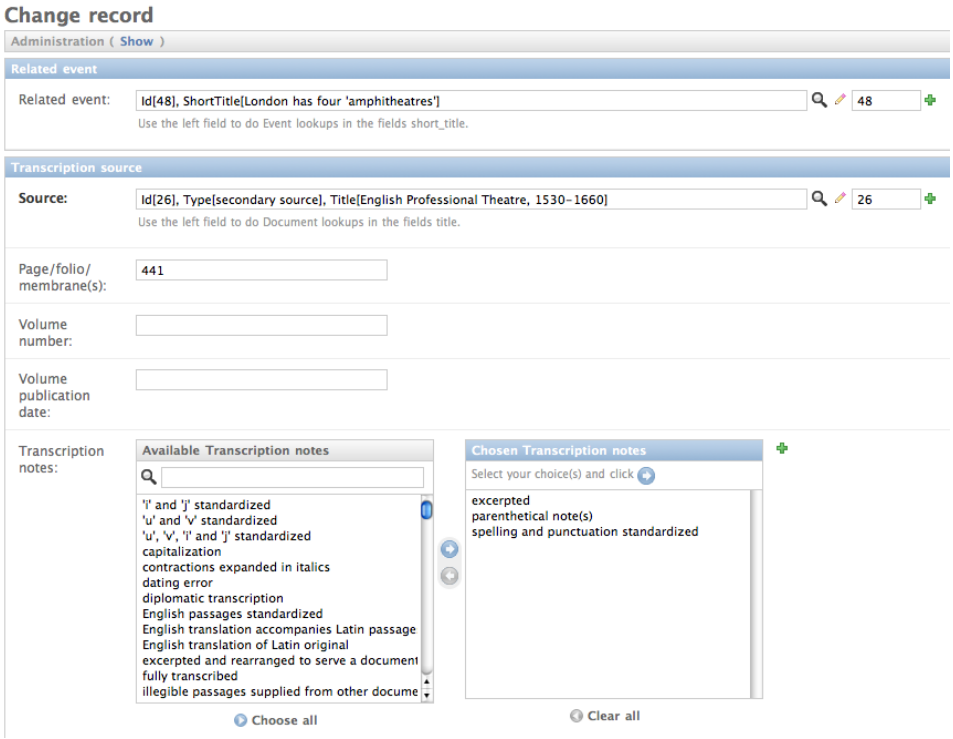

Figure 4. EMLoT administrative web application for the "record" entity

3.2 Designing a data model: a broader perspective

To sum up what has been said: when designing a data model the first crucial thing we need to observe is, clearly, *what* we want to represent. Secondly, a major constraint is given by the specific *medium* we are using to store our data. We have seen that in order to take advantage of the characteristics of a relational database, data must be organized into tables that are related to each other (furthermore, even if we haven't discussed this aspect, we must remember that often data get organized in a specific manner because of performance issues). Finally, there's a third principle playing an increasingly important role in the modeling of data: the *problem of interoperability*. With this notion we refer to

the fact that, ideally, we would like to be able to integrate other people's databases with the least effort, and similarly, we would like other people or institutions to be able to reuse our research results.

This interoperability problem is quickly gaining importance because of a recent phenomenon that is happening on the web. If we look at the evolution of the Internet, it is easy to realize that if at the beginning the web was conceived mostly for human usage and consumption, now the scenario is quite different. This is due to a number of factors, including the increase of computing power, the availability of cheaper and larger storage devices, and last but not least, the constant growing number of internet users. As a result, more and more *structured* data sources (like databases or XML files) are being made available online, and, consequently, we are now facing an emerging *web of shared data* that is way too vast only for people to make sense of. Connecting the dots of this enlarged network requires more sophisticated approaches than the current ones, since such new approaches must enable a "deeper" interlinking of related resources. These recent developments of the web are happening in various forms, which we will not discuss here,[7] but the key aspect is that computer programs can orchestrate such sources of structured data very efficiently.

For example, consider this type of scenario: a student, after finding out about a theatrical company through EMLoT, wants to search the REED collections for other materials about this company, and see these results displayed within EMLoT. Or maybe our student would like to seek more information about the patrons associated with this company by querying the *Patrons and Performances* website. However, this is just the tip of the iceberg. There are many other data sources out there, providing, for example, relevant information that focuses on the more geographical, historical or artistic sides of the subject we are investigating. These "lateral steps" are in principle doable right now, but they require a lot of copying and pasting, changing sites, and performing different searches which might just make us lose track of the context from which we departed. So, the key issue here is how to connect these resources at a "deep" level, so as to allow a

[7] For an introduction see Berners-Lee et al 1999; Bizer et al 2009.

more seamless integration? If each one of them has been created using a different but overlapping data-model, how can we guarantee that we can make the resources "talk" to each other?

3.3 The ontological approach to data modeling

Clearly the interoperability problem is equally as important as it is difficult to solve. What we want to highlight here is an approach that will not solve the problem by itself but that can help us in organizing our data so that interoperability is facilitated. This is called the *ontological* approach to modeling and in general it can be seen as a useful best practice for modeling data that can be "consumed" not just by our application, but also by others, thanks to the infrastructure provided by the web.

This approach comes from the discipline of ontology engineering (Mizoguchi. 2003; Schreiber. 2007), a research area in artificial intelligence that devised a way to employ a rich body of theory from philosophical ontology to the purpose of making conceptual distinctions in a systematic and coherent manner. Ontology engineering is concerned with making representational choices that capture the relevant distinctions of a domain at the highest level of abstraction while still being as clear as possible about the meanings of terms. The term "ontology" is borrowed from philosophy, where ontology is a systematic account of existence. In computer systems, what "exists" is exactly *that which can be represented*.

In order to give the reader a short introduction to this approach in the following sections we describe one of its fundamental principles, and then explain how it has been applied to the context of EMLoT.[8]

The principle can be summarized as follows: we must determine an essential property for each concept and instance in the model, and make sure that this property is correctly *inherited* in our hierarchies. The notion of inheritance here has a technical meaning: it refers to the fact that in a given hierarchy, if a super-class (e.g., "animal") has

---

[8] For a more comprehensive description see, for example, Guarino and Welty. 2002

a property (e.g., “is mortal”) then all of its sub-classes (e.g., “human”) must have that property too. Furthermore, in ontology we say that a property of an entity is not essential if it just happens to be true of it, accidentally, for all time. For example, for a sponge “being hard” is not an essential property, although some sponges can be hard. Instead, for a hammer, “being hard” is essential as we cannot think of a hammer that is not hard. So, if we have a hierarchy of concepts in the model, the principle tells us that we must make sure that this essential property is inherited, otherwise our model is likely to generate inconsistencies. In what follows the same principle is used to clarify the relationship between the concept of “human” and “teacher”:

> Let us take the common example: <teacher is-a human>. Assume John is a teacher of a School. Given the usual semantics of is-a, since John is an instance of teacher then he is also an instance of human at the same time. When he quits being a teacher, he cannot be an instance of teacher so that you need to delete the instance-of link between John and teacher. However, you have to restore an instance-of link between John and human, otherwise John dies. If we are only interested in property inheritance between human and teacher, the relation <teacher is-a human> seems to be valid because any teacher is a human in any case. However, if we think of essential property and/or identity criterion of classes, then we can understand the relation is inappropriate and would cause such a problem. (Mizoguchi et al. 2007)

This example applies quite well to our initial EMLoT model: we do not have “teachers” in our domain, but we have “authors,” “editors” and “historical individuals” whose essential property is different from the one of “person.” This can be easily motivated by the fact that an “author” doesn’t cease to be a “person” when she decides to change her career. Nonetheless, we modeled “authors” as sub-concepts of “person.” It is important to

highlight that our approach may work in the restricted context of EMLoT—mainly because our database (at this stage) does not contain information about people who start and stop being authors—in other words, that is because we do not have to cope with alternative possible states of reality.

However, when integrating our database with others we might have to represent this information too. So, a more solid way to organize the "person" concept would be needed. This can be achieved by adding a "role" concept that lets our person-instances be authors, editors, or anything else without generating any contradiction (see Figure 5).

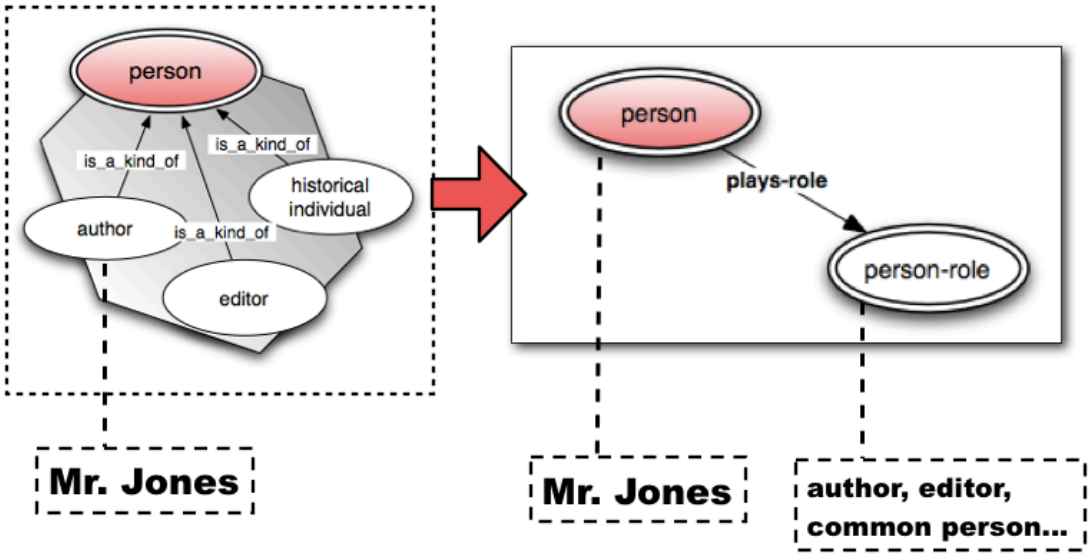

Figure 5. Applying the identity principle to EMLoT model

In conclusion, the ontological approach provides us with a way to create conceptual models that, being deeply rooted in the "shape" of reality itself, are much more solid. This is especially useful when we face the task of putting together (i.e., integrating) multiple overlapping views of the same reality (i.e., databases).

Figure 6 shows a better version of EMLoT data model, which has been created by applying the ontological approach more extensively. Notice that also the "source" concept hierarchy has been improved: in this case, by applying the "identity principle"

we got rid of the "original" and "transcription" concepts and transformed them into relations employed by the "record" concept when referencing sources.

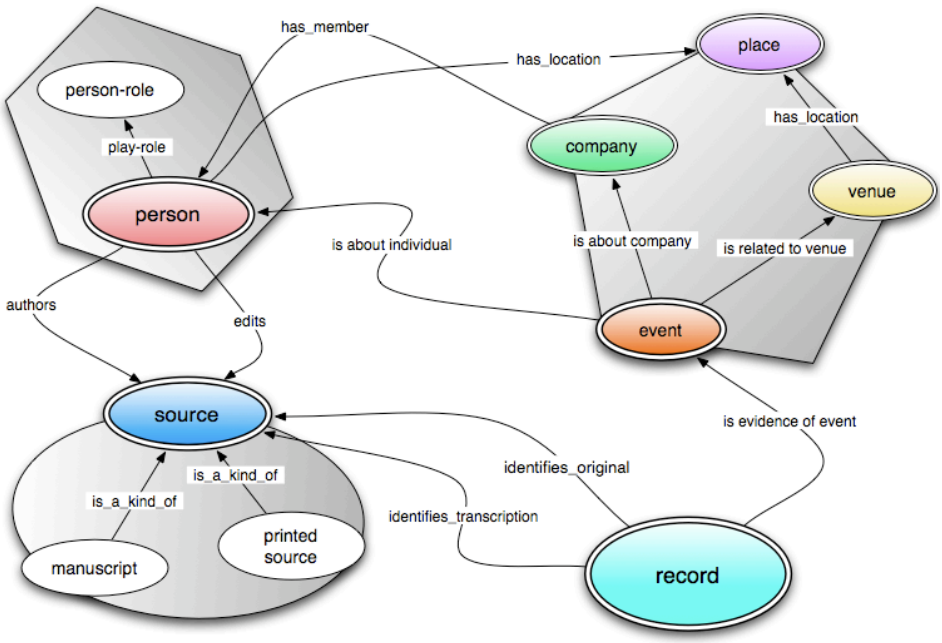

Figure 6. A better model for EMLoT in the light of the ontological approach

4. Conclusions and future work

In this essay we introduced the context and purpose of the *Early Modern London Theatres* project by situating it within the pluri-decennial research activities of the Record of Early English Drama project. In particular, we discussed in details the approach used in creating the database and the type of information that it contains. We highlighted the fact that EMLoT stands out among other related resources for it addresses simultaneously two research needs. First, it is an extensive bibliography, insofar as it aims to identify, record and assess transcriptions from primary source materials relating to the early London stage, as found in secondary source print and manuscript documents. Second, it provides a comprehensive historiographical survey of the field, as it stores information

about the people, places and happenings that emerge from reading the aforementioned transcriptions.

For these reasons we expect that the EMLoT web application will meet the needs of a variety of scholars, and stimulate the interest of many non-specialists too. In order to support further the latter type of audience we also made available a "learning area" section (http://www.emlot.kcl.ac.uk/learning-zone) that compensates for the highly specialized character of much of the information in the database with more gentle introductions to the field, video lectures, and other learning materials created by our team of experts.[9]

In general, it is fair to say that the database modeling approaches described in this article were successful in representing the meaning of the information we are considering with a high degree of accuracy and precision. Nonetheless, due to the necessarily circumscribed context of this resource and the practical purpose of the project, in some cases the transformation of real-world descriptions of (aspects of) theatre history into more formal computer representations forced us to adopt 'workarounds'. That is, to opt for suboptimal modeling solutions that can act as 'working approximations' of particularly complex aspects of the portion of reality we are examining. For example, the fact that *writers* or *actors,* within the EMLoT data model, would be better represented not as *types* of people*,* but as *roles* that people can play within specific contexts (cf. section 3.2). This kind of simplifications are not unusual for digital resources creators, especially in cases where, like in EMLoT, the nature of the domain being represented is particularly intricate and semantically rich.

To the purpose of providing a more generic and comprehensive formal representation of the 'world' emerging from the EMLoT project, we have started working on the expansion and refinement of the database model herewith presented so that is becomes a full-fledged formal ontology. In particular, we argued that this type of activity is of primary importance if we decide to consider our database work within a broader context,

---

[9] We are fortunate to have received a 2010 SSHRC Public Outreach grant to enable the development of the Learning Zone feature.

that is, the context emerging from the fact that more and more databases similar or tangential in scope to EMLoT are being made available online. This scenario, often referred to in the use of terms such as *semantic web* or *web of data,* calls for an infrastructure that supports an increased level of interoperability between databases. For example, in a fully developed web of data we could easily query different repositories by using a common interface, and then republish these results elsewhere so that other people can use it. We have introduced the *ontological* approach to conceptual modeling, a technique that supports the creation of more solid conceptual models, and shown how such an approach can be applied in EMLoT context for the purpose of facilitating any future data integration task. We are currently finalizing the first version of an ontology that models all the major entities in the theatre history domain, so that separate digital resources, such as EMLoT and the various other *REED* online materials, can be accessed simultaneously.

These results will be made available in a separate publication later this year. In the meanwhile, it is our hope that this publication will contribute to raising the awareness of the importance of such topics for the purpose of creating a 'digital ecosystem' of online resources centered on theatrical studies; we therefore invite other digital humanists to get in touch with us to the purpose of creating a special interest group.

References

Berners-Lee, Tim, M Fischetti, and T.M. Dertouzos. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web By Its Inventor* (San Francisco: Harper).

Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. **"**Linked Data - The Story So Far**,"** *International Journal on Semantic Web and Information Systems - Special issue on Linked Data***:** 1-22.

Guarino, N., and C. Welty. 2002. "Evaluating Ontological Decisions With Ontoclean," *Communications of the ACM* 45.2: 61--65.

MacLean, Sally-Beth, and J.A.B. Somerset. 'Performers on the Road: Tracking Their Tours with the REED Patrons and Performances Web Site,' in *New Technologies in the Renaissance*, William R. Bowen and Raymond G. Siemens, ed. 2008. Medieval and Renaissance Texts and Studies 324 (Tempe, Arizona: Iter Inc./ ACMRS), 39-51.

Mizoguchi, Riichiro, E. Sunagawa et al. 2007. "A Model of Roles Within an Ontology Development Tool: Hozo," *Journal of Applied Ontology* 2.2: 159-79.

Mizoguchi, Riichiro. 2003. "Tutorial on Ontological Engineering - Part 1: Introduction to Ontological Engineering," *New Generation Computing* 21.4: 365–84.

Schreiber, Guus. 2007. "Knowledge Engineering," *Handbook of Knowledge Representation*, ed. Frank van Harmelen et al (Elsevier Science), 929–46.

Tvarožek, M. and M. Bieliková. 2007. "Personalized Faceted Browsing for Digital Libraries." In *Research and Advanced Technology for Digital Libraries*, edited by László Kovács, Norbert Fuhr and Carlo Meghini, 485-88. Lecture Notes in Computer Science, Vol. 4675/2007 (Springer Berlin / Heidelberg).

Wickham, Glynne, William Ingram, and Herbert Berry, eds. 2000. *English Professional Theatre, 1530-1660* (Cambridge: Cambridge UP), 654.