Building a modern manuscript and archive editing platform: the EMAN experience

Richard Walter

Journal of Data Mining and Digital Humanities

Building a modern manuscript and archive editing platform: the EMAN experience

Richard Walter [1*]

1 Institute of Modern Texts and Manuscripts (ITEM) CNRS / Ecole normale supérieure

*Corresponding author: Richard Walter richard.walter@ens.fr

Abstract

EMAN (Manuscript and Digital Archive Publishing) is a digital publishing platform for the dissemination and exploitation of modern manuscripts and archival holdings. It supports the management of a digital publishing project as soon as a document is online, including metadata production, data relationships, transcription, publication and data value enhancement, etc. It includes many functionalities based on the Omeka software and complies with the open formats of the web. Its objective is to be a truly multiple purpose publishing tool for corpus dissemination projects in the humanities. Our article describes the methodology used to build this tool in a situation where few examples were available to draw on.

keywords

Corpus, digital publishing, genetic editing, project management, OMEKA

INTRODUCTION

EMAN (Manuscript and Digital Archive Publishing) [1]is a digital publishing platform for the dissemination and exploitation of modern manuscripts and archival holdings. It has been available since spring 2015. In spring 2019, the EMAN platform covered 29 projects as varied as the digital edition of Thresors de la Renaissance, press corpora or ENS course packs, genetic sources for operas or comics, the genetic edition of a tale by Paul Valéry, Vittorio Alfieri's biography, Gaspard Monge's or Emile Zola's correspondences, or genetic studies on 19th century Franco-Chinese dictionaries. The projects published range from a genetic "micro-edition" based on a very small corpus to the mass publication of very varied documents, with or without their transcription in the case of some projects. Historically, the coverage spans from the Renaissance to the 20th century. Many languages are present in digitized and edited documents, and the typology of media is also very broad, from manuscripts to videos.

We decided not to start creating an ad-hoc platform from scratch, because this would have involved heavy development and uncertain maintenance. We chose to rely on the open source

[1] http://eman-archives.org

ISSN 2416-5999, an open-access journal

Omeka software,[2] a tool that is highly appreciated in the digital humanities community for publishing corpora or scholarly collections. We adapted this tool and implemented editing methodologies that we will discuss in detail in this article.

The editorial workflow and navigation interface of our Omeka-based EMAN platform have undergone a number of specific developments to adapt them to documentary and scientific practices on modern manuscripts and archival collections. This is how we are building a digital publishing model. But every tool is the result of its own history and context. Both shaped the technical and methodological choices we made to implement this platform.

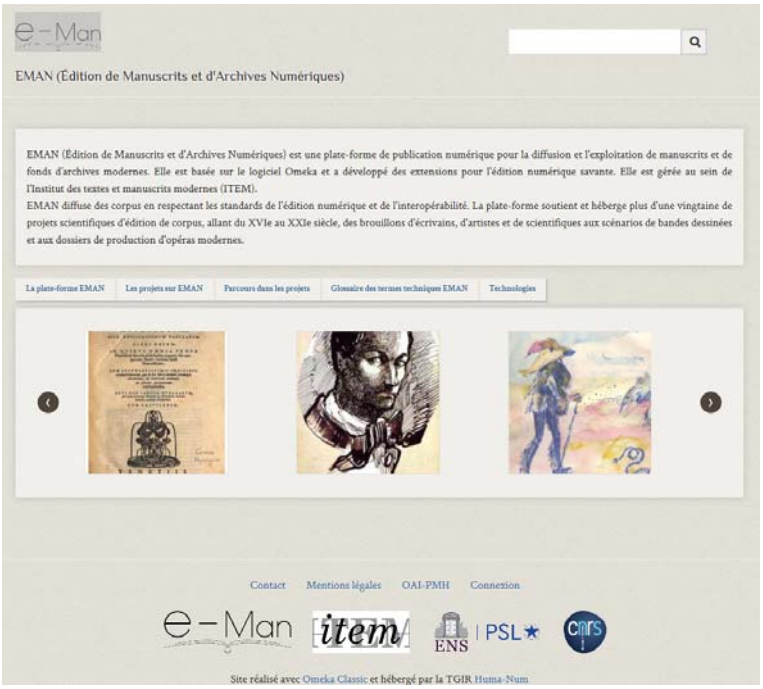


Figure 1: home page of the EMAN platform
(www.eman-archives.org)

I BACKGROUND OBSERVATION

With the new context of widespread digitization, research and publishing practices are undergoing rapid change. A broad-scale transition to digital applications is underway. We are shifting into a new paradigm: one can photograph a document with a mobile phone, produce documents faster, manage increasing amounts of data, share one's discoveries, or be rewarded earlier, etc. Another change is that now one can expect to find and exploit multidimensional and exponentially growing corpora. It is possible to process both text and images and there is even an interest in hybrid, semiotically complex objects growing up. On top of that, we no longer experience the pressure and finiteness of publication in a fixed object. But, there are high risks of data and knowledge loss. Can one still remember the data production methods used in one's penultimate project, and which were its criteria and technological or editorial blocks?

Digitizing and editing a document requires taking into account its material (documentary and archival) dimension and its history (genetic dimension), with multiple types of information on its content and also in its comments. These diverse pieces of information must be structured but the standards for their description follow different rules. At each stage of the digital project, editors face different kinds of choices depending on the kind of publishing planned: scholarly

[2] https://omeka.org

publication *versus* technical constraints; the interoperability necessary for sharing *versus* specific technological options, according to one's own needs. There is a certain amount of awareness needed to realize that the edited digital object itself is transformed by the editing process, which is particularly true with digitization, as it results in a change of the initial document.

Digital publishing is subject to the same requirement and quality principles as "classic" printed publishing, with some additional technical imperatives, such as compliance with digital standards. *However*, it facilitates updates, opens up new perspectives and a depth of analysis unhindered by the cost and finiteness of the paper edition. It is crucial to have a structure for digital information to achieve this online edition, and both for its dissemination and archiving. It is therefore necessary to deal with the norms, standards, methods and customs of the digital world. As a result, new - not necessarily all digital - tasks emerge before, during and after the act of publication. Digital management requires above all a very fine-tuned upstream preparation, as the choice of a tool or method can have irreversible consequences.

II PLATFORM REQUIREMENTS

In this context, various digital publishing projects[3] came together in the ITEM laboratory to choose a common solution to respond to their needs to massively and simply publish archival scientific corpora in a sustainable way. In the summer of 2014, we benchmarked existing tools and techniques and began blueprinting the EMAN platform as a digital publishing tool for the distribution and exploitation of modern manuscripts and archival collections.

The different options available and implications brought changes to the platform, but its objective remained twofold, to ensure dissemination and exploitation. Its working material is wide: from the manuscript to the archive collection. Currently, its only restriction is to impose a coverage limited to the modern age of printing. A specificity of the platform is to draw attention to the progress from a draft to the published object, or failing this, to provide a genetic approach to the document creation processes.

Therefore, our purpose is not to occasionally edit a few texts with their scholarly apparatus. Our first presumption was not to focus on a definition of the manuscript or archive but to accept all documents used as sources to produce the document studied or to ground a scholarly analysis. Our definition of the source is deliberately very open so we can deal with as many special cases as possible. This is why we opted for large, and even very large, volumes. We manage hefty volumes of highly diversified content, considering them from four different angles:

1. The scanned documents;
2. Their documentary and genetic descriptions;
3. Their transcripts and scientific analyses;
4. The relationships between the different contents from a genetic perspective.

Our first specification was and remains "to keep it simple", that is, to publish images, texts and various digital sources online with an editorial workflow that meets the scientific requirements of a critical and genetic edition and enables the exploitation of the disseminated content as material for an analysis of creative processes.

[3] The FLIM and Lettres françaises projects quickly joined the CHispa and Manuscrits francophones which all had one thing in common: their members did not have a digital editor and computer scientist on their team. Other projects joined within ITEM or as part of partnerships with other institutions.

So, we naturally decided not to develop our own tool, which would require costly and technologically burdensome developments, and above all unsustainable maintenance. We are well aware that, in the context of digital projects, demands and requirements change with advancing developments and technical responses to initial problems. We wanted to keep a strong scalability and reactivity to the proposed tool.

Finally, it is imperative to ensure the sustainability and interoperability of the data needed to exploit and extend the value of research outputs. The tool must fulfil the numerous functions that match these four levels of consideration for the documents that are to be edited on the platform:

1. Edit digitized documents with metadata that comply with international norms and standards;
2. Associate these documents with one another through multiple logical, temporal and genetic links (by creating a genetic file);
3. Link the digitized document to its enhanced transcription;
4. Ensure interoperability for archiving, dissemination and "harvesting" to make data open to the semantic web.

In addition, this platform had to be generic and modular to adjust to publishing projects that differed very much in terms of history, corpus size and the diversity of scientific questions addressed. The tool had to propose or even impose a straightforward and multiple-purpose structure, corresponding to digital publishing standards, and also bricks matching the needs of each project (not all of them require transcription or geolocation). Therefore, it is crucial that one could use all the tools a project needs as a common basis.

III WHY OMEKA

During our benchmarking phase in 2014, we analyzed the possibilities of relational databases which could be compelling for data analysis (e.g. dBase or Fedora) but required too much customization for an online publishing system. We also reviewed Content Management Systems, such as WordPress, Drupal, MediaWiki or bibliographic managers like Zotero, which are very efficient to meet specific needs (news and bibliography) but the editorial approach of which was too rudimentary for scholarly publishing. We looked at the online image sharing tools (e.g. Flickr) that can be used to manage large masses of images, while only offering minimal image metadata management features. Finally, we discussed with colleagues similarly sophisticated editing projects on heritage corpora, among others with the designers of the Telma[4] and BVH[5] platforms, who work on the highly specialized scholarly edition of many texts using XML-TEI files. Our resources did not enable us to undertake such projects immediately and we also needed to create a digital library based on our many digitized image repositories.

This is why we selected the Omeka software. This web publishing system specialized in online museum collections, digital libraries and scholarly editions is at the crossroads between databases, content management, publishing and digital document enhancement systems. It is being developed by the Roy Rosenzweig Center for History and New Media[6] at George Mason University in Virginia, USA, which already developed the Zotero bibliographic software.[7] The goal of the creators of this easy-to-use tool was to help users to "focus on content and

[4] http://www.cn-telma.fr
[5] http://www.bvh.univ-tours.fr
[6] https://rrchnm.org
[7] https://www.zotero.org

ISSN 2416-5999, an open-access journal

interpretation rather than programming". In the digital humanities arena, this free software is a fast growing success. Its code is open and the application is easy to install on a web server, so it effectively makes up for the scant computer skills and resources of many research teams.

Omeka therefore makes it possible to publish content in a basic and, above all, very flexible way. Its underlying architecture remains straightforward and can be adapted according to the needs of each project by adding modules (plug-ins) and using visual models (templates). In contrast, the Dublin Core scheme[8] is a required to structure metadata with this software to maintain sustainability and interoperability. The tool therefore meets our requirements for data continuity and interoperability, which are necessary to process and enhance the corpora and archival collections we intend to publish. What remained to be done was to adapt this tool to our needs, and not our needs to the tool.

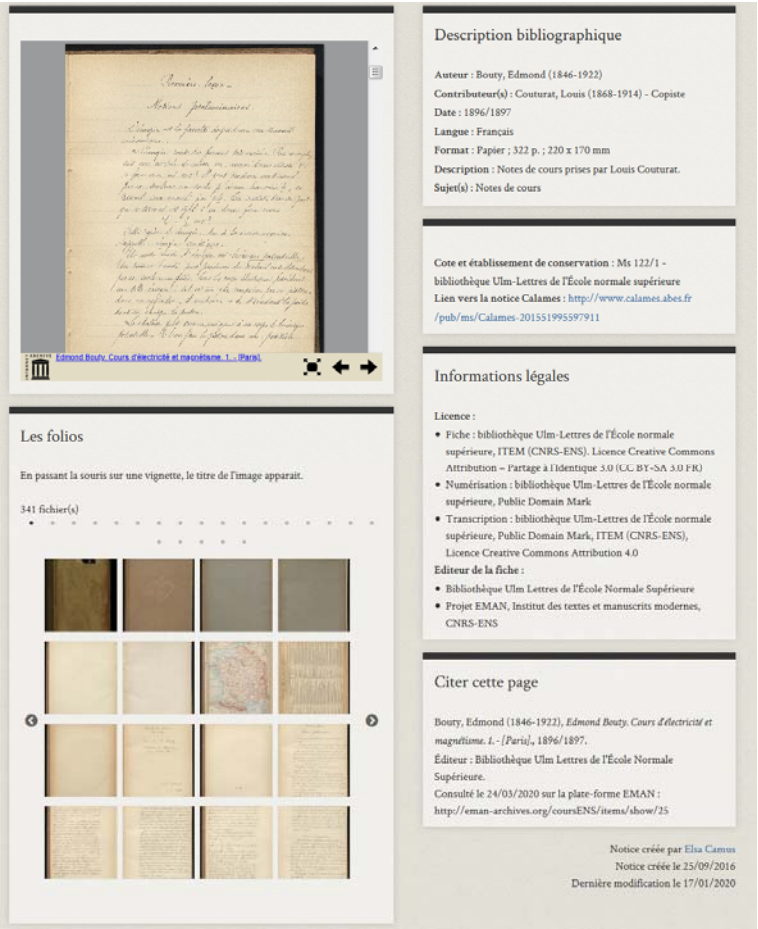


Figure 2: *Note de cours de l'ENS* project sheet with images and DC metadata (www.eman-archives.org/coursENS)

IV A SCHOLARLY AND DIGITAL EDITORIAL APPROACH

Consequently, we started to build a publishing tool based on Omeka. We had a strong basic prerequisite in mind: digital publishing should not be considered as a simple digital reproduction of a paper document. There is a need for intellectual and editorial thinking that leads to the act of digitization or digital publishing. And digital publishing must reflect this process.

[8] http://dublincore.org

In our view, document digital encoding is to be construed as a process of understanding and not as a limited vision of the work. Our assumption is that a digital edition is mainly used to trace the life of a document, from its genesis to its distribution. It is therefore necessary to publish its genetic file containing all the sources that led to this document in addition to the reproduction of the relevant document. Here, the digital conversion of sources and textual variants of the published work (or the version considered final) changes its perception and requires new methodologies which remain to be theorized in this digital context.

Is our approach a genetic one? The contribution of digital technology to theory in genetic criticism leads to quantitative and qualitative changes in the analysis of creative processes. With our platform, we want to produce a methodology for the digital editing of sources that can be used to understand any creative process. This specificity is reflected in the content written in the metadata of a document or collection and in the structuring of the data itself or the creation of navigation and visualization tools as well. During each of these stages, we focus on the creation processes, the circulation of documents between one another as a whole. It is then necessary to carry out work to structure and develop a set of interconnected data. This work is based on the notion of a "genetic file" for a document and must be integrated into the digital production flow of its edition.

The development of a genetic file relies on the relationships which need to be established between the different stages of the creative process. One of the technical objectives of a digital variant will be to mark out links between these steps to better show the ongoing processes. The philological and genetic interest of the digital technology lies in the fact that it is possible to combine the edition of successive versions of the document in the same virtual space, but also to exhibit the whole material of the avant-textes (aka foretext or pre-text) that falls under the activity of writing and invention like variants and metatext (reminders, personal notes, drawings, etc.).

It is therefore necessary to adapt an editorial approach that takes this genetic dimension into account and matches the particular background of a platform supporting multiple projects. These projects differ substantially in their depths of analysis and objectives. We publish projects in parallel on the EMAN platform when they cannot be arranged by subjects or chronologically. Our context is that of a virtual library which gathers all its documents in its collection and on which we can use the same genetic or documentary analysis framework.

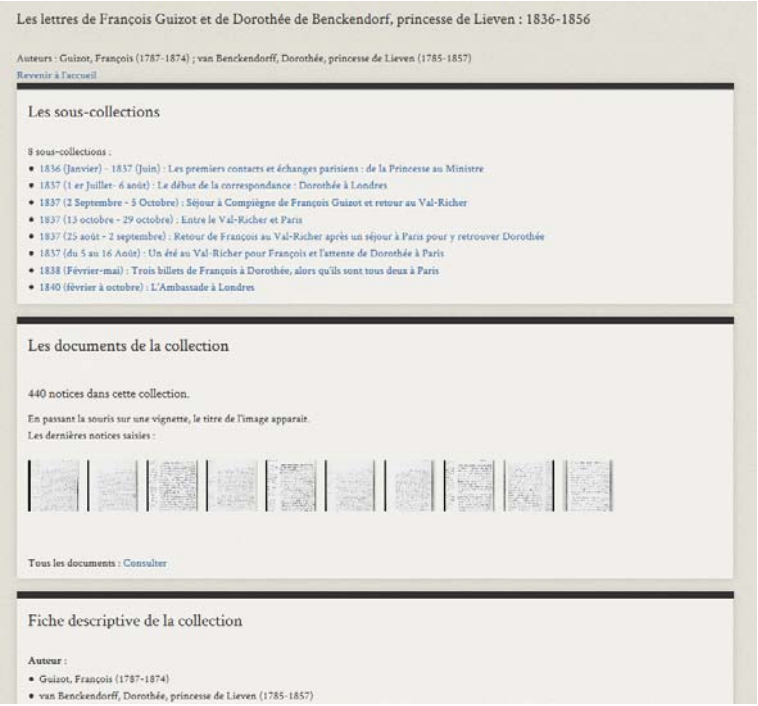


Figure 3: presentation sheet of a collection from the *Guizot-Lieven* project (www.eman-archives.org/Guizot-Lieven)

V A NEW EDITORIAL APPROACH

This diversity of projects requires that we take a flexible and iterative editorial approach and propose a platform that meets the needs of project leaders and expected users. Above all, it is necessary to take into account two aspects in the position of a researcher who is involved in a digital editing project and in which he/she has to take on two functions at the same time:

1. Make his/her scientific focus accessible to other researchers and a wider public, in a form that matches its characteristics and the researcher's requirements;
2. Interpret the data collected and published, identify recurrences and induce regularities which are specific to the work, its author, genre and time.

The digital context will extend the number of solutions available to meet this dual requirement. But one must not become entangled with the management of this new abundance of opportunities. For example, one of the observed limitations of a digital edition is the time length and complexity of the coding modes of the different elements. One could be encoding forever and *ever*. Where and when should one stop doing this?

We made a radical and rather new choice in the context of traditional scholarly publishing, which involves publishing a scholarly rich object with all the necessary footnotes. We bypass the difficulty of coming to a refined edition of some content. This bears the risk that it will never be published on line, while this is a context providing the opportunity to update one's work at any moment. We start by editing the document with the minimum basic encoding so that it can be read, processed and posted in a digital archive. Then we provide the tools for more complex edits by adding description, relationship and analysis features as the research work and the setting up of the tooling on the platform move forward. Our project management approach is really to work step-by-step and edit results in successive layers.

In fact, managing such bulk of documents is more challenging than ever considering the massive growth of digital content, the "variability" of document statuses and reading and research needs. What is more, no "turnkey" method or solution is available for the moment. All requirements apply at the same time: as soon as someone accesses a document, the data must be of high quality, well-edited, well-transcribed and well-translated. However, there is a great need to follow a phased approach imposing a time line to the way of solving different issues as one problem solved gives rise to another and one advancing research highlights another aspect long considered secondary, and so on.

We secure the publication and durability of the data as a minimum at first on the EMAN platform, never giving the impression of providing a finished scholarly product. In this digital context, one tends less and less to believe in finiteness. This is why we opted for Omeka in which there is a requirement to have the documents indexed first with the Dublin Core standard, which was adopted as the norm for document exchange internationally. It may seem simplistic as a way of describing documents as complex as writers' manuscripts or videos of modern operas, but in reality Omeka only imposes fifteen fields to be able to work better afterwards. With Omeka, researchers / publishers can add descriptive or analytical metadata according to their projects. That is the possibility which made Omeka successful, notwithstanding that of customizing one's data editing and exploitation by including tools and processing features.

Customizable metadata is therefore an additional layer to Dublin Core metadata. It can be extended indefinitely and makes it possible to enrich the description as needed in the project, even at the risk of creating a maze of fields in the record. However, one can sort this metadata by types (document description, administrative rights, interpretations), thus facilitating comparisons or exchanges with other projects. This also reflects a paradigm shift, since we no longer move from the exhaustive description of the document towards its simplification to provide it with indexing, but rather build digital foundations to be able to deploy a more complex architecture.

The first and crucial step to establish these foundations, and thus to type the metadata, is the data modelling phase. This is an important moment in project management on EMAN. This modelling does not occur from scratch. It is based on the scholarly requirements of the project researchers, which are analyzed in comparison with the digital editorial experience acquired within the platform. There is a need to tap on documentary and archival insights to secure the successful conformation of data sets and their digital insertion into EMAN.

Therefore, we have to think about a model for processing and displaying all types of data that can be integrated into the project, at the start or a later stage (which is even harder to figure out). Thus, we have to choose the right granularity, i.e. the basic documentary unit in the project. Additional questions must be addressed. What will the basic record refer to in Omeka's structure? What will be the criteria used to gather records into a collection? What are the contents one should assign to the very broad definitions of Dublin Core metadata? A whole host of questions must be addressed before even formatting the data entry and visualization forms.

As opposed to what one might imagine, this step does not take place on screen but involves going through a process of discussion, confrontation and controversy. An acute mania for meetings can be a significant risk at this stage of the project and at some point it will be necessary to walk the talk to dive and work into the tide of data. One necessity must be repeated over and over again: "we must return to the data" during these multiple discussions. However, as digital a project can be, it must go through an initial and crucial phase of experimenting on paper and around a table, very far from "all-digital" designing.

Figure 4: structure of collections and records in the *Valéry-Robinson* project (www.eman-archives.org/Valery-Robinson)

VI TOOLS & EXPERIMENTATION

We use this common approach in all platform projects and until their completion. The time to integrate the data into the EMAN platform comes when the modelling phase has reached its end. Here too, we provide a system to support the editing and use of data in accordance with the objectives of the platform. It was not invented from scratch but was built gradually through experience gained with the projects arriving on the platform. It relies on the installation of one Omeka instance per project and includes a specific EMAN theme and plug-ins, which aggregate all the modifications and additions we derived from Omeka software, but without modifying its core.

The editorial model consists of data structured into collections and sub-collections, with Dublin-Core descriptions and customized metadata types, tools for virtual exhibitions and for transcription into TEI format (Text Encoding Initiative),[9] tools for CSV-spreadsheet[10] or XML-file[11] imports and exports of all or part of the corpus. This system is complemented by numerous guides for recommendations and data entry, tutorials and weekly collaborative workshops.

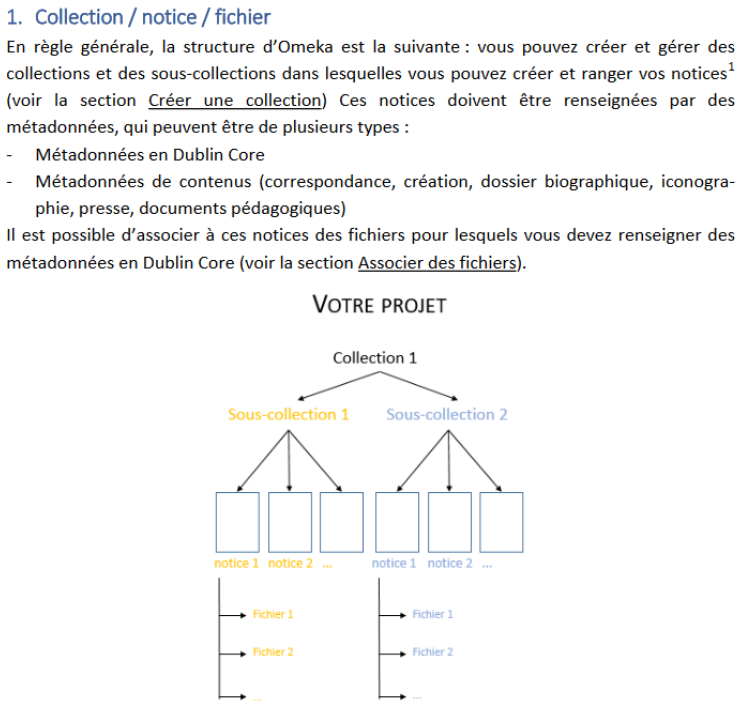
Project members must take up and adapt this environment according to the characteristics of their corpora and/or the aims of their research. The model can change according to the decisions of the platform's steering committee. It is worked on during workshops and study days. All adaptations and manipulations are stored in a collaborative - and restricted - space and a research notebook[12] is used to disseminate the editorial and scientific activity produced with the platform. Once again, the management of this editorial process is significantly time intensive and slows the process of public data publishing. However, this process, albeit slow, also fosters the emergence and organization of a user community who shares more than the mere use of a common tool.

We continue to enhance the tools and procedures provided and create new ones. There is a constant need to adapt to the documentary, research or technical practices of those who lead the projects hosted on the platform. New projects are coming in and especially as more expectations

[9] https://tei-c.org
[10] Comma-separated values : https://en.wikipedia.org/wiki/Comma-separated_values
[11] Extensible Markup Language ; https://en.wikipedia.org/wiki/XML
[12] http://eman.hypotheses.org

are met, the exploitation of the results produced leads to new needs. All tools designed in a platform project, or funded as part of it, are made accessible to all others and their development is always discussed collectively among EMAN teams. Moreover, the advantage of a platform like Omeka lies in the opportunity to join a wide community of developers. We make our developments for EMAN available to the Omeka international community as extensions. The EMAN page of the GitHub code-sharing platform,[13] provides ten plug-ins to date, including those allowing users to customize the display of a collection, a record or a file for each project. This is how we can rely on venues of exchange where codes or techniques are available and thus continue to improve and extend these functionalities.

While keeping our approach experimental, we designed an editorial methodology which is specific to EMAN. It is in keeping with traditional project management techniques and supports with particular emphasis interactions between research and editorialization, dissemination and experimentation, constrained research and heritage needs, etc. Also, note that updated structuring and support documentation that is understandable by everyone is necessary, e.g. a charter of good practices, numerous workshop and study day reports, various tutorials and guides to data preparation and entry into EMAN, all of which, through multiple editions, become a real methodological encyclopaedia of digital publishing.


1. Collection / notice / fichier

En règle générale, la structure d'Omeka est la suivante : vous pouvez créer et gérer des collections et des sous-collections dans lesquelles vous pouvez créer et ranger vos notices[1] (voir la section Créer une collection) Ces notices doivent être renseignées par des métadonnées, qui peuvent être de plusieurs types :

- Métadonnées en Dublin Core
- Métadonnées de contenus (correspondance, création, dossier biographique, iconographie, presse, documents pédagogiques)

Il est possible d'associer à ces notices des fichiers pour lesquels vous devez renseigner des métadonnées en Dublin Core (voir la section Associer des fichiers).

Figure 5: page extracted from the input guide for writers on EMAN

VII THE TRANSCRIPTION AND VISUALIZATION PROJECTS

With 29 hosted projects and about 100 active "data entry operators", we have reached an interesting critical mass of people to share problems and resources, to address them and pool the management of projects. One should not consider the multidisciplinary and transversal characteristics of approaches around EMAN as a risk of dispersion. On the contrary, by working on common elements, we are building more effective methods for dissemination and sustainability - while not preventing in-depth work specific to each approach.

[13] https://github.com/ENS-ITEM.

There were multiple initial constraints to make this object adaptable to the wide variety of experience and skills among the participants in the different projects. A maximum array of forms and screens had to be available, as well as the vocabulary used. Conversely, it was also necessary to avoid using Omeka's technically sophisticated tools offering a maximum number of functions but without a graphical interface. We chose to provide all possible action in WYSIWYG mode to facilitate learning the Omeka working principles and EMAN tools, so that tasks are performed through onscreen forms and without having to ingest technical codes. Note that the people who have to enter content into the platform are not "professional encoders" and often even come from communities unfamiliar with such technical facilities.

Two ongoing projects on transcription and visualization illustrate our EMAN methodology. We started them to meet scientific needs and to make these compatible with the EMAN method and platform. Moreover, even if specific programmes exist, like Oxygen or Gephi, their user interfaces are made for professionals, and there are not any devices available yet to adapt them to generic platforms, massive data and for a non-specialist user base.

Our editorial project now aims to provide users with the opportunity to view text transcripts next to the images of published documents, especially those that are the most difficult to decipher. But the perspectives are changing. The usual problems of transcription, i.e. proposing exhaustive diplomatic editions, reporting all phenomena as seen (or perceived?) on the document, take on a new significance with Omeka or with the various forms of digital publishing platforms. Omeka gives access to the image of the document and the user has the possibility to view the transcription next to this image. Thus, he/she can now see the text apparatus while reading it in another format. So, isn't there a considerable waste of time encoding phenomena that can be viewed on screen? We no longer need to reprint but rather give access to massive texts available for searches and working out hypotheses.

To propose transcripts or rather transcription levels on EMAN, we combined the software's capabilities, in particular the user friendliness of its procedures, to the encoding capabilities of the Text Encoding Initiative. In other words, we wanted to accommodate the huge library of TEI tags in a WYSIWYG entry mode.

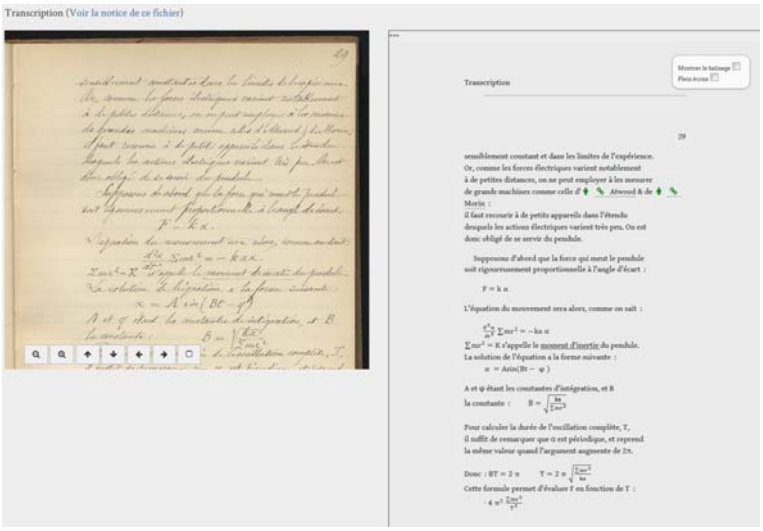


Figure 6: transcribed page of the *Note de cours de l'ENS* project (www.eman-archives.org/coursENS)

We kept the same approach for other tooling issues by using exclusively one standard (but not necessarily entirely) and by customizing one tool to meet the needs of a specific community.

We worked on integrating a toolbar into our entry forms to encode text in TEI using add-on buttons that can be customized to the needs of a project. We keep here to a word-processor-like feel: the user selects text, then the icon of the desired encoding, he/she sets this encoding in a window, validates to enter the tags. This helps get around learning computer language but also- and above all - avoids the risk of mis-typing tags. The result is TEI-encoded text that can then be exported to an XML/TEI file for other uses elsewhere. We have developed a Transcript module to handle this WYSIWYG encoding on a very wide range of phenomena. This module is not as powerful as a specialized XML/TEI encoding software, but it meets most of the transcription needs of platform users. If they want to do more extensive encoding, they can export their EMAN results to specialized software, such as Oxygen, which is used in many projects for text encoding.

There is another project which is located further downstream on the document editing process. We are working on visualizing the relationships of a document within a set. Our aim is to show the relationships that can exist between all the variants of a text grouped into a collection. To obtain this, one would need a typology of relationships so that they are not only chronological or alphabetical. Hierarchical and cause-and-effect relationships are not self-evident in artistic creation processes. The representation of creative processes cannot be unambiguous, as elements can be found in several places in the same process. It is then necessary for us to invent relationship models and test them on our corpus.

We are thus working to make relationships available as proposed by various semantic web initiatives. Our assumption would be to be able to specify certain types of relationships leading to schemas tracing influences, evolutions, and circulation between data. This can be referred to as "multi-layered" relationships. We have adopted graph or mind mapping techniques to display them, with the difficulty that many of them are based on a driving element that serves as a starting point for visualizing relationships: establishing the starting point of a work in a stemma codicum may not be straightforward. However, the combinatorial power of the encodings produced by Omeka helps to obtain visualizations from several points of view or even according to the fields of analysis of several interpretative approaches. For the moment, we have a module that we have designed, Graph, by using a VisJS script library and adapting queries to the specificities of our corpora. This tool can display all the relationships in document or between documents for the same collection in the form of a static graph. From there, we hope to be able to obtain dynamic graph visualizations based on customizable criteria.

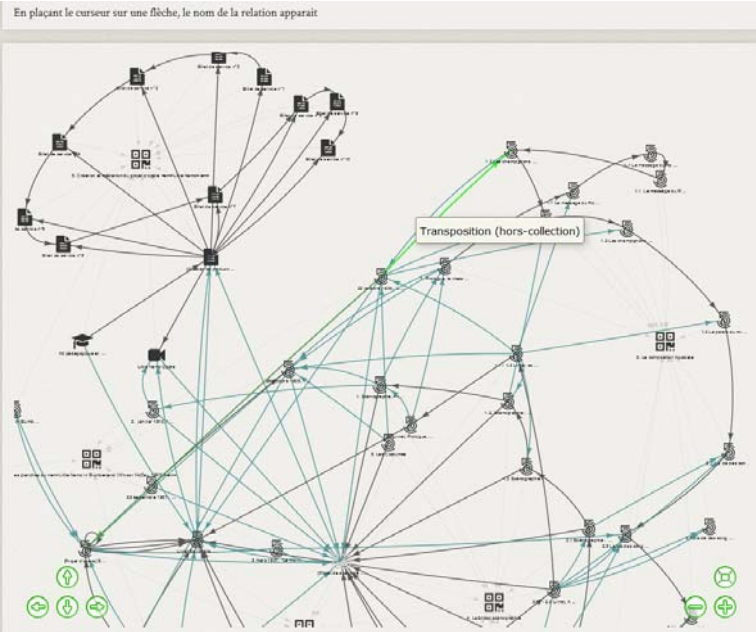


Figure 7:visualization of the relationships between the records of the *Les Sources génétiques de l'opéra* project (eman.hypotheses.org/127)

CONCLUSIONS

Digital applications make it possible to publish and manage heterogeneous data in a single virtual space. A document can be published in different forms (raw or corrected transcription, formatted or not, displays showing sources or relationships, etc.). The publication space is unhindered by volume or character size limitations. We shifted to new scales (or even a new time period?). With digital technology, we no longer have the same obstacles to publishing, we can even multiply dissemination modes through elements converging from different media (sound, videos, images, etc.) and by representing the different existing levels of relationships between documents. Technically, we can publish all the versions of the same text without limitation, with their transcriptions together with comments and analyses and with the genetic file that informs about the creative process. But one must accept that this publication only represents a current state of research and that it is bound to change with the latest findings or different points of view. This is actually what multiple editing is all about.

We intend the EMAN platform to be a representative of this type of edition. We built it on the potential of Omeka, a robust generic tool used by a large community. We refined certain functions of this tool to meet the specific needs of our user community. Features like display customization, transcription or graph display had not been planned by the Omeka's creators. Our developments are made available to the international Omeka community as extensions on the code sharing platform GitHub with about ten plug-ins released to date. This openness makes it possible to share and improve our experience, but also to perpetuate and disseminate our work by facilitating its adoption in other similar projects.

References

Cavalié E., Clavert F., Legendre O. et Martin D. (dir.). *Expérimenter les humanités numériques. Des outils individuels aux projets collectifs*. Les Presses de l'Université de Montréal (Montréal), 2017.

Charnois T., Legallois D. « Que faire des corpus (une fois) numérisés ? L'exemple du discours littéraire », Université Caen/ MRSH, 2013. http://www.unicaen.fr/recherche/mrsh/document_numerique/959.

Claerr T. Westeel I. (dir). *Manuel de constitution de bibliothèque numériques*, éditions du Cercle de la Librairie (Paris), 2013.

Claerr T. Westeel I. (dir). *Manuel de la numérisation*, éditions du Cercle de la Librairie (Paris), 2013.

Dacos M., Mounier P., *L'édition électronique*, La Découverte (Paris), 2010.

Mounier P. (dir.). *Read/Write Book 2 : Une introduction aux humanités numériques*, OpenEdition Press (Marseille) 2012.

Rouffineau G. *Éditions off-line, Projet critique de publications numériques 1989-2001*, éditions B42 (Paris), 2018.