Engaging the User: The *Encyclopedia of Diderot and d'Alembert: Collaborative Translation Project* and New Scholarly Paradigms

Kevin Hawkins (University of Michigan, Ann Arbor)
Jason Kuznicki (Johns Hopkins University)

Kevin speaks:

- The views expressed here are my personal opinions. Nothing is an official position of the Scholarly Publishing Office.

- We summarize our mission like this: "The University Library, through its Scholarly Publishing Office, is committed to providing academic publishing services that are responsive to the needs of both producers and users, that foster a sustainable economic model for academic publishing, and that support institutional control of intellectual assets. The Scholarly Publishing Office seeks to disseminate high-quality, cost-effective scholarly content through both print and electronic means."

- Most publications are freely available online.
- Authors usually retain copyright to their works.
- We publish many journals and a few monographs. We see the Web as our primary medium, but we also do print versions for some works. We've only done print monographs, though, since we don't have a good procedure at this point for producing monographs easily without a lot of design work.
- This shouldn't be confused with one of the other functions of our office: to handle the ordering of hard-copy reprints of digitally reformatted items from the Library's print collection, but this doesn't involve the production of new content, which is generally how we distinguish ourselves from DLPS, our better known cousin at Michigan. So while this has gotten a lot of attention, we don't see it as our primary work.)

- I'll talk about one publication of ours that is neither a monograph nor journal. The *Encyclopedia of Diderot and d'Alembert: Collaborative Translation Project* is building an online, freely available English-language edition of the *Encyclopédie*, originally published from 1751 to 1777 and containing more than 70,000 articles.

- The project directors wanted to make this monumental piece of scholarship available to a wider audience, particularly undergraduate students, so non-specialists, who don't necessarily speak French, could

gain study the *Encyclopédie* as a historical source, gaining a better understanding of Enlightenment intellectual history.

- Only a handful of articles in the *Encyclopédie* had ever been translated, and many of these translations are in out-of-print books or books that were too expensive to ask students to buy just for the useful section.

- The project was begun by clearing copyright permission for many of these previously published translations and publishing these, and by adding a few new translations by the project directors.

- It has steadily grown to 196 translations online as of last week.

- The project directors assign articles to volunteer translators based on their expressed interests. Submissions are reviewed by the project directors and sent to us as RTFs. We convert to SGML. Anyone can volunteer provided they have a sufficient command of English and French, but we ask that undergraduate students have a faculty member review their translations.

- We deliver the project, as almost all of our other publications, using the DLXS Middleware with the XPAT search engine. Our texts are in DLXS Text Class, meaning they use a modified version of TEI Lite that's even more flexible than TEI Lite in its content model. We do something like Level 3 encoding according to the TEI Text Encoding in Libraries: Guidelines to Best Practices.

- I'll turn things over to Jason Kuznicki, a colleague from Johns Hopkins who I cajoled into giving you the perspective of a contributor and user of the project.

At every conference I have ever attended, there has always been at least one speaker who, when he began his talk, has caused the audience to sit back in their chairs and quietly ask themselves, "What *was* he thinking?" When I accepted the invitation to speak before the Digital Library Federation, I had the sneaking suspicion that I might be that person. And, having listened to three surpassingly technical presentations already, I now have no doubt whatsoever: I *will* be that person. And you may very well ask yourselves: "What *was* he thinking?"

I am not a computer scientist, or a programmer, or a systems administrator. I'm not even a librarian. Compared to all of you, I am probably the least computer-literate person in the room, if not in the entire building. I am here to give a user's perspective on the digital media in the humanities, focusing on the *Encyclopédie* Collaborative Translation Project, on which I've done some work in the past. I would like to begin my remarks with a quotation that is very familiar in my own field and may well be in yours as well. The great French historian Emmanuel Le Roy Ladurie declared in 1979 that, "Tomorrow's historian will have to be able to program a computer in order to survive."

Twenty-five years later, I *am* tomorrow's historian, and I still cannot program a computer. In one sense, it's not surprising: Nowadays we distinguish between programmers and content providers, a distinction that was by no means clear in 1979. But as the gap between the two has widened, however, it has become more and more important to allow for effective communication between the programmers and the content providers. In the big picture, I fear that historians have been quite reluctant to become content providers in the digital media, and I believe that their reluctance is partly because of this growing divide. Only very recently and tentatively have

historians embraced the digital media, and they have done so quite slowly even when compared to other disciplines in the social sciences and the humanities.

The issues and subject matter that historians examine have changed tremendously since the nineteenth century, but our methods of production and dissemination have more or less stagnated. In the last hundred years, we have created entirely new fields of history, including demographic history, labor history, the new economic and cultural histories, and the histories of gender and sexuality. But production and dissemination remain very similar to the model perfected with the rise of the modern research university in the nineteenth century: Historians gather information at phyiscal archives; they digest information at universities; they lecture on it to students, and they write up texts with formats that are substantially similar to those produced in the nineteenth century. That is to say, these texts are almost always articles or books that are published in paper-based, peer-reviewed outlets.

Where computers have entered the discipline, they certainly have not revolutionized it—at least not yet. Structurally, the computer has replaced the typewriter, leaving the rest of the process more or less intact. Historians may take laptops into the archives, but notetaking and information transfer methods are frequently just digital re-creations of the old-fashioned paper notecard. I can recall that as an entering graduate student, I asked how I might better use computers in gathering my information. And—at a university that is *not* Johns Hopkins—I was told that I shouldn't bother: A tenured faculty member and a well-known authority in the field declared in no uncertain terms that notecards were the best and really the only way to do research: Write the cards in pencil while you're at the archives. Then come home, set the cards out on your floor, and arrange them into the story. To this day, it remains the way that many of us work.

As a tool for dissemination, we view the Internet with suspicion. Do a search for the Holocaust, and you will understand immediately why this is. Generally speaking, what history can be found on the web ranges from the mediocre to the mendacious. The peer review process has created very high standards in most published historical journals. But it's powerless to weed out

the Internet cranks, and let's face it, there are a terrible lot of cranks on the Internet. Conspiracy theorists, promoters of racial hatred, plagiarists, and advocates of thinly-evidenced "alternate" histories abound. In one term, a student submitted a paper that had been plagiarized from a Holocaust-denial website; the following term, I entirely forbade my students from using the Internet, and promised to fail them if they did. Even before a group such as this, I can't honestly say that I regret the decision.

For those of us who know how to find them, however, there are certainly some bright spots out there, and the *Encyclopédie* Collaborative Translation Project is undoubtedly one of them. The original *Encyclopédie* is arguably the most important document produced in eighteenth-century France; if anything, its importance is surpassed only by the Declaration of the Rights of Man. While not the first encyclopedia, it was remarkable in its systematic approach, its wide scope, and its enunciation of the Enlightenment idea that all things are in principle knowable. The eighteenth century was the first age to share that ideal; it may well have been the last age where human knowledge was still small enough to render the project of compiling it all possible. This compilation is exactly what Diderot envisioned, and to that end, he engaged some of the brightest minds of his era: Besides himself and d'Alembert, the *Encyclopédie* featured articles by Rameau, Montesquieu, Buffon, Turgot, and Voltaire. In today's terms, this would be like arranging a collaboration between Stephen Hawking, Madeline Albright, Richard Rorty, J. D. Salinger, and Andrew Lloyd Weber.

Historically, the *Encyclopédie* is the starting point for virtually all questions touching on the Enlightenment, as it set the tone for how an entire generation of philosophers approached the world. The work's direct descendants, from the *Encyclopedia Britannica* all the way to Wikipedia, owe the *Encyclopédie* an enormous debt. And to speak still more generally, the project of the *Encyclopédie* is also the project of the modern library itself: The systematic classification of human knowledge is an idea that we now take for granted in large part because of the efforts of

Diderot and d'Alembert. Thus the *Encyclopédie*'s importance to the historian is absolutely paramount.

But how are we as historians to teach this source? At 70,000 articles, very few have ever read the entire thing, and I'm certainly no exception. Making it accessible to a non-francophone general reading public virtually requires any presentation of the *Encyclopédie* to be rapidly browsed, scanned for interesting content, and parsed according the needs of the user. Only with the rise of the computer has such a project truly become feasible. The original format of the *Encyclopédie*, which was arranged not alphabetically, but by the branches of general knowledge as understood in the 18th century, is cumbersome and difficult even for specialists like myself to use. Within the past few years, most of us have come to rely on French-language electronic versions of this text in part for this very reason.

The Collaborative Translation Project now features nearly two hundred articles in translation, and these offer an excellent starting point for non-francophone students to learn about the state of 18th-century knowledge in many—though certainly not all—different fields. One of the most important factors determining the relative success of the project so far is that the *Encyclopédie* Collaborative Translation Project does not exist in a vacuum. One reason that historians trust the veracity of the translations so far is simply the good reputation of the contributors, a reputation that has mostly been built in more conventional channels. For example, Dena Goodman, was already a leading published historian of the French Enlightenment long before she took over supervising this project. The same is true of many others who have contributed articles; indeed, as an advanced graduate student rather than a full-time professor, I am a rather junior partner in the whole enterprise. The Collaborative Translation Project, then, represents an expansion of traditional academic activities, but it is certainly not yet a fundamental revisioning of how historians do their work.

Besides the sources of the project's authority, much of its *material* also comes from more conventional channels. Most articles that have been translated so far have come at the prompting

of requests made on the Internet mailing list H-France, which is a private, highly-selective list open only to graduate students and faculty; that same list is a frequent discussion site for articles in progress and those that have already been translated. Indeed, it is hard to imagine the project ever succeeding without the H-France mailing list to supplement it. While individuals not associated with the list may certainly submit translations, my impression is that they seldom do. To my knowledge, the production of new translations has not been assigned in history classes; for most undergraduates, the work is beyond their abilities. French language classes would probably be a better venue for the assignment, and I understand that it has already been offered on several occasions, and I look forward to this type of practical and often quite fascinating assignment becoming more common in the future.

Still, the extant translations were for the most part done by specialists in their own subfields of history. For example, I have translated several articles from the subfield of religious history, where I could be expected to give a more competent translation than most. On the other hand, I have left the mathematics entirely alone, because I must confess that I simply don't follow it. One weakness of the Collaborative Translation Project to date is that many of the translated articles reflect the interests and areas of relative expertise of those who have done the translations. Historians have done the great bulk of the work so far, and we tend to be well-versed in those fields most closely aligned with our own; this bias is reflected in the selection of articles so far, which tends to favor material on economics, anthropology, religion, languages, and other disciplines in the humanities and social sciences. The sections on engineering and mathematics have not been so well-represented in our efforts to date, although they are quite numerous in the original text. One way of correcting this bias would be to enlist bilingual individuals with competence in these fields to do some translations as well.

One last way that the project might be improved is ambitious, but I thought I would offer it anyway, simply to generate discussion and to perhaps learn a bit more about the technical limitations of the project itself. Just as the Collaborative Translation Project does not exist in a

vacuum, so too one text does not make history—not even a tremendously important text like the *Encyclopédie.* Even before it had fully appeared, contemporaries began critiquing, correcting, and in some cases flatly condemning its content. Voltaire devoted much of the last period of his life to a series of "Questions about the *Encyclopédie,*" often writing point-by-point refutations of others' articles. And the debate has continued ever since.

Indeed, one of the best definitions of history that I have ever heard is that *history is an argument about the past.* For the Collaborative Translation Project to really do the work of history, besides merely being an excellent resource, it would have to have a much greater measure of user feedback than it currently possesses. Besides the translated article, could contributors perhaps be invited to submit commentaries on their translations, lists for further reading, notes on difficult words or phrases, and relevant historical responses? Could we perhaps implement a discussion forum under each article, much as weblogs now have?

I do not know whether these steps are technically feasible or even whether the principal architects of the project would approve of them. From my perspective, though, I could see how all of these steps could make the *Encyclopédie* Collaborative Translation Project a more useful teaching tool. In a larger sense, they might also bring more of the work of history to be done online, and we can only hope it might day replace the cranks. So to start the discussion, the biggest question I have might actually be too general: I'd like to know how much of what I have just said was new or surprising to you, how much of it was familiar, and how the technology of both this project and more generally the digital approach to history can be made better from our interactions. While historians are adept at providing content, very often we don't at all understand the available systems of collection and dissemination on which our careers may increasingly depend.

Kevin speaks again:

- Library as publisher – an experiment, but we're already finding ourselves suffering some signs of institutionalism in terms of our technology. We're reluctant to break with our established methods not just because they're comfortable but also in order to maintain consistency and interoperability. Naturally we're also limited by time and money.

- Currently we do not have cross-references activated as hyperlinks, but we hope to have this feature up soon. We will also have a translated diagram of the tree of knowledge available soon.

- The first analogy that comes to many people's minds when I explain this project is Wikipedia. A wiki would provide a great framework for this project and allow truly collaborative translation and editing, but we have not moved in this direction because it would require our office to maintain another delivery system and require using another search system separate from those of our other publications. We would also lose the ability to include the project in a cross-collection search, such as we're able to do now when all collections are in DLXS with XPAT. Just as important to us as providing more a more author-friendly copyright system and low barriers to access is ensuring long-term accessibility and preservation. I don't need to explain the costs of digital preservation and concerns over the long-term obsolescence of content and architecture standards, so I'm sure you can understand why we try to keep our publications as standardized as possible and all using the same delivery system. However, we find ourselves doing quite a bit of customization as it is, and it looks like we will have a huge amount of programming work to do in the near future when DLXS release 12 comes out, which will have major architectural changes in it. So I, at least, am starting to question to whether ostensibly using a standard like SGML and only one delivery system is really saving us trouble in the long run.

- A more attainable goal would be to encourage our authors to adopt Creative Commons-type licenses, or simply force all contributors to publications to use a certain license or "a certain license or better". There's nothing stopping us from this except figuring out a consistent way to store rights metadata in our file headers (not absolutely necessary) and to display these rights to users.

- We would like to develop a more automated content transformation process.

- We plan on sometime implementing DOIs and Unicode compliance.

- Maybe we'll have institutional repository interoperability one day.

- If you're interested in hearing about our recent cooperation with the ARTFL Encyclopédie Project (which presents the entire French original online through a subscription model) to create a bilingual controlled vocabulary for use by both projects, please speak to me afterwards.