

Report of the John Donne Society's Digital Text Project (JDSPP)
December, 2018ⁱ

Brent Nelson, Project Director
University of Saskatchewan

I. Assessment of the current state of the Donne Prose project

The project was born at the twenty-fourth annual meeting of the John Donne Society in Baton Rouge, 19-21 February 2009, when the executive committee of the society adopted the John Donne Digital Prose Project and named me its director. The originating idea, posed by Gary Stringer, was to build a prose archive that would complement the Donne Variorum's Digital Donne site, which was almost exclusively dedicated to poetry. Implicit was the ultimate goal of merging the two resources to achieve a complete archive for Donne's entire oeuvre. After a year or so of consultation (including a survey of the society) and planning, I started parcelling out the main work of the first phase of the project: preparation of transcriptions. By the time a fleshed-out prospectus was published in 2013, the project was well underway. As context for an assessment of how far we have come in seven or so years of labour, I must first summarize the plan as laid out in the 2013 article that set the agenda for Phase One of the project.

Phase One Prospectus: a summary

It was imagined that a full digital prose archive would aim to bring the prose works of Donne to the same state of preparation as those of the Donne Variorum's *Digital Donne*. It would therefore comprise a complete set of transcriptions of every witness of every prose work by Donne. There were two main aspects to this work:

1. fresh digitization of exemplars (original documents)
2. completion and correction of EEBO-TCP transcription against these exemplars using XML markup following the "TEI-light" guidelines. (In this last point, the prose works would differ from the Digital Donne poetry texts, which at that time did not use the TEI-XML standard for markup.)

In this prospectus, we set out a few principles and elements that would guide the development of these materials:

1. The archive will be open-access.
2. It will adhere to the high standards of quality and accuracy that scholars have come to expect from their textual resources, subject to rigorous peer review (where available), editorial oversight, and accountability to an advisory board. We did not, however,

establish an editorial board, for reasons which will be discussed below. In our directions for transcription work, we now define our standards with reference to the MLA Guidelines for Editors of Scholarly Editions (See “About the Project,” Donne Prose wiki).

3. It will acknowledge the nature and extent of every participant’s contribution.
4. It will be rich in meta-data.
5. It will plug into available scholarly infrastructures to make the archive is visible and easily accessible.
6. It will be group-sourced, built by a community of interested users, beginning with the John Donne Society.
7. In phase one, it will be designed to go forward without major funding.
8. It will support other editorial projects and, in turn, will explore ways in which those projects might contribute to the archive.

Phase I: Progress Report

As noted above, in our prospectus, we identified three aspects to the work of phase one: fresh digitization of exemplars (original documents); completion and correction of EEBO-TCP transcription against these exemplars; and XML markup of the transcriptions using a simplification of the “TEI-light” guidelines. To these, I have since added a fourth: development of a data structure to enable navigation of and access to the primary texts.

Digitization of exemplars:

One of our first steps toward producing high quality, accurate transcriptions was to acquire and work from high quality, clear images, i.e. ones that are significantly better than those provided in EEBO. We now have fresh images for at least one witness of each of Donne’s prose works, and in some cases multiple witnesses, amounting to probably half of the material we need to collect (by page count). We have acquired our images by five main avenues:

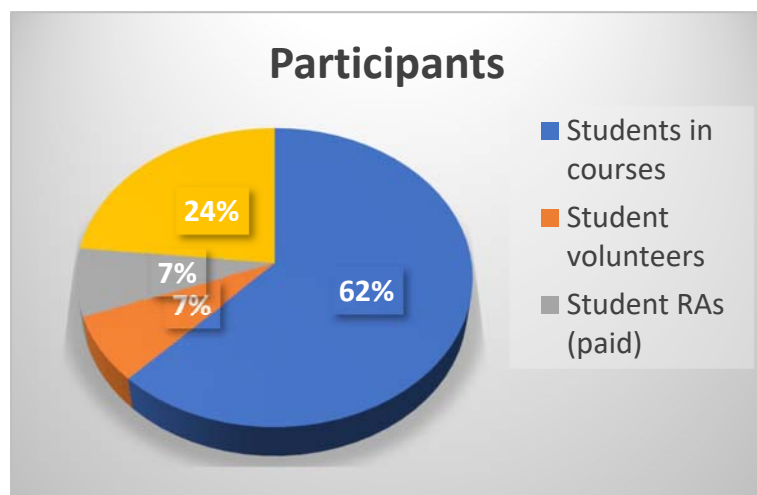
1. images from the Donne Variorum project;
2. a few images from print facsimiles, which we plan to replace with original images;
3. images acquired by Donne Society members of materials held in their own libraries;
4. images requested directly from institutions where desired materials are held;
5. images acquired by asking our libraries to purchase or borrow materials and then digitizing them (chiefly at the University of Saskatchewan).

These images are, for now, being published to the blog as pdf files. So far, we have focused on images that were relatively easily to obtain. The latter half of our digitizing efforts, reserved for Phase Two, will be more difficult and might require financial resources.

Production of transcriptions:

Transcriptions are being produced through two methods: correcting EEBO-TCP transcriptions, which is the focus of Phase One; and producing new transcriptions of manuscript witnesses, chiefly reserved for Phase Two, but also occurring in Phase One. As noted above, most of this work is being achieved through volunteer, “group-sourced” labour, beginning with members of the society. In 2013, we had some thirty-three participants who were making or had made some kind of contribution to the project, mostly in proofing, correcting, and augmenting EEBO-TCP transcriptions. About two thirds of these were faculty members of the society, and the other third were graduate students or undergraduates, most of whom were employed by me or another society member with modest funding from our home institutions. Over time, other society members and associates became involved in different ways: making requests to their own libraries to have primary materials digitized; incorporating transcription work into senior undergraduate and graduate courses; processing secondary sources for eventual inclusion (work that was outside our original Phase One plan). Graduate students participating in our annual conference became interested in the project and, for a time, became one of the most active group of volunteers contributing to the transcription work. The other increasing sector were students involved in courses taught by society members who made transcription work part of their course activities.

To this point (December 2018), we count 124 past and continuing contributors:



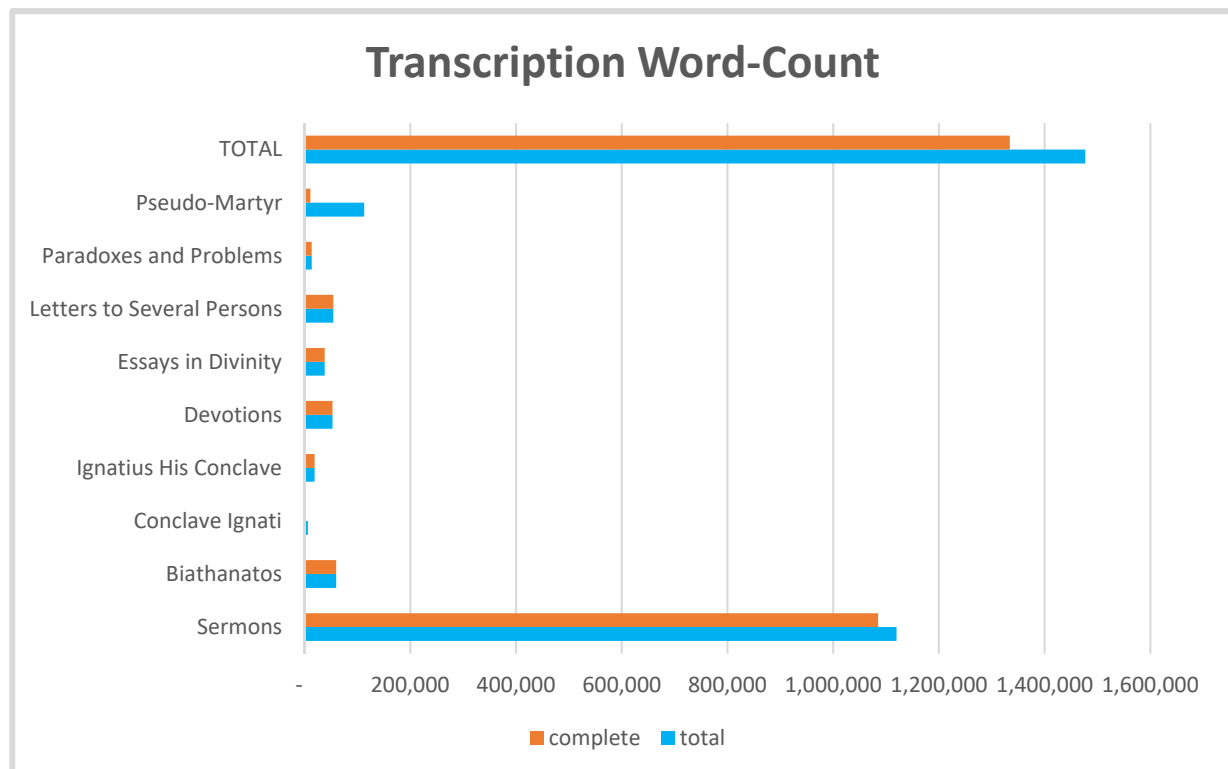
29 faculty or non-student volunteers
(including one non-academic)

77 students involved in course work

9 paid student research assistants

9 student volunteers

We have made good progress, although there has been some deviation from the original work plan. We intended to focus on the sermons and complete them first before moving on to other works, but for a variety of reasons (some of them volunteer preferences, some of them curricular), we added other works into our mix along with the sermons. Another major alteration is that for Phase One we no longer plan to produce a complete set of transcriptions of all witnesses, for reasons I will explain shortly. Instead, our plan for Phase One is to produce at least one transcription for at least one witness of each of Donne's works. By this standard, then, we measure our progress thus, by word-count:



Some of these transcriptions are ready and published on our blog (donneprose.blogspot.ca/), while others require more post-checking processing (tweaking and adding of XML markup, etc.) before they are made publicly available, but our intention is to make all our materials available as they are ready.

II. Lessons learned from Phase One thus far

1. Group-sourcing can be both inefficient and effective.

We confirm what other large-scale transcription projects have found, that volunteer participation is most efficient when participants commit to completing a significant volume of work: for

efficiency sake, it is better to have a few super-contributors producing several transcriptions than several contributors producing one or two transcriptions each. The ideal is to have volunteers who have been trained and are able to continue producing work with very little supervision. Effectiveness, however, cannot always be reduced to mere efficiency. The benefit of involving a large number of participants, as we have done, is that more people feel a sense of ownership and investment in the project. This approach is also beneficial in training students (see further below). We believe that in the long-term, there is much to be gained by this inefficiency.

2. Group momentum wanes.

In our 2013 prospectus, we recognized that the progress of the project would be largely determined by enthusiasm in the society as its central motivating force. If enthusiasm is measured by participation in the hands-on-work of the project, I think it is fair to say enthusiasm is waning. Those most inclined or able to contribute to the work of transcription checking as volunteers have already contributed significantly, and I am reluctant to keep going back to the same group asking for more. For the project to meet our revised objective for Phase One, we need more volunteers, preferably ones who will commit to completing all or a major portion of one of Donne's remaining larger works: *Biathanatos*, *Pseudo-Martyr*, or *Conclave Ignati*.

3. Motivation can be found in convergence.

One area where new interest has been generated is in faculty incorporating transcription work in their teaching at both the graduate and undergraduate level. To echo point one, while inefficient from an administrative perspective, involving students in course work is rewarding for the students and contributes to our over-all mission as teacher-scholars of training and equipping students. Many of the students really enjoy the work and feel gratified in contributing to a real-world project rather than a merely academic exercise; and it gives them hands-on experience in understanding how texts are produced and transmitted. It is not, however, an efficient way to accomplish the work, because it involves a great deal of preparation and supervision for a limited term of involvement.

4. Success breeds administrative challenges.

Administration of the work has been a challenge. As noted above, as a means of managing the work and ensuring the quality of the materials we produce, we imagined we would establish an editorial board. Our intention was to have a number of editors involved to help supervise the work, but finding others to join me has proven a challenge. One needs to be at a certain stage of career at a certain kind of institution to commit to this kind of work. As a result, I have been the only editor distributing and managing the work, creating a bottleneck when participation and

production levels are high. Exacerbating this bottleneck have been some problems with the software we use (which has been in development through the course of Phase One of the project), requiring additional editorial intervention and interrupting the work-flow of volunteers.

5. Success breeds motivation.

This point is speculative: it might be the case that slowed momentum is owing, in part, to the lack of a full-function website. Our intermediate plan was simply to provide links to materials, as they were ready, via our project blog: <http://donneprose.blogspot.ca/p/materials.html>. The hope was to have a full set of transcriptions in place before developing a proper, fully searchable and functional website. One remedy for waning enthusiasm, it seems to me, is to give our stakeholders—chiefly, Donne scholars—an interim resource that begins to deliver on the promise of what the ultimate fruit of our labour will be. At present, for example, because our sermon transcriptions are presented as printed collections (such as the *LXXX Sermons* [1640] or *Fifty Sermons* [1640]), it is difficult for a user to locate a particular individual sermon of interest. Moreover, our set of texts is incomplete: we do not have an accessible corrected text for each work. For this reason, I am proposing a revision to our Phase One work plan.

III. Looking forward: next steps in Phase One and beyond

Revision of our work-plan for Phase One

Based on lessons learned, it is apparent that we need to adjust the scope of our Phase One plan to make sure we have something that is complete in a way that is useable for scholars (a text for every work), and then expandable in scope (eventually to include every witness of every work).

- a. *A revised schedule for preparation of materials.* In order to have a basic, working site as soon as available, it seems sensible to reduce the scope of Phase One: we will now aim to produce a transcription of a single witness for every work. A complete set of transcriptions, covering every witness of every work, will wait for Phase Two. At the same time, because our one area of continuing participant enthusiasm is in the context of course content, we have to be open to instructors electing to work on documents (witnesses) that are not part of this plan.
- b. *Publication of uncorrected transcriptions.* Ideally, we would not want to publish materials that don't meet our quality criteria outlined above, being contrary to one of our founding principles. One might also wonder whether having a full set of texts available, however imperfect, might be enough for some of our stakeholders and thus reduce motivation to push

our work forward. I think, rather, that if we publish a transcription that is clearly and prominently marked as “Uncorrected,” we might draw attention to the need to do this work, and thus draw the interested user to take up the task. Having a complete set of transcriptions, now, will also enable us to produce a workable site that is searchable and navigable across Donne’s entire oeuvre.

- c. *A basic site for navigating materials.* Another deviation from the original Phase One workplan has been development of secondary resources, again arising out of the interests and curricular activities of participating members. Some of these will be part of Phase Two (bibliographic resources on scholarly publications of criticism on Donne), but another set of resources (descriptive bibliography of the primary materials) are also being processed to serve as the backbone for navigating our Phase One texts. These resources also fulfill one of our objectives outlined in the Phase One prospectus: that the archive will be rich in meta-data.

Phase One website

The two standard bibliographic resources for primary materials related to Donne are Peter Beal’s *Catalogue of Literary Manuscripts 1450-1700* (CELM: <http://www.celm-ms.org.uk/>), which includes information on all manuscripts containing works by Donne, save for a handful of recently discovered manuscripts, and Geoffrey Keynes’s *A Bibliography of Dr. John Donne*, which provides detailed bibliographical information on each of Donne’s printed works up to the twentieth century. The first of these has been copied from the website, parsed, and ingested into a relational database. The second of these resources has been OCR’d from print, proofed, and parsed, in preparation for ingestion into the same database. To begin, this content will stay in the background and be used only for linking content, but if we are able to acquire permission from OUP, we would like to be able to offer at least part of the Keynes to our users as metadata. CELM, which appears to be offered in the spirit of open-access, is already available for use, but ultimately, we would like to approach John Lavagnino about collaborating to extend our adaptation of CELM to include the poetry, and perhaps work toward a complete ingestion of the site into database form. The key functionality afforded by these resources is navigation by work—including, in our case, individual sermons, paradoxes, and problems—and by source document (manuscript or print edition). These bibliographic resources, along with STC numbers, library shelfmarks, and the sigla and short forms developed by the Donne Variorum for Donne’s works, manuscripts, and print sources, provide us with nomenclature for identifying and relating

each *work*, *document* containing a work, and *copy* of each print document, correlated with each *text* (transcription) we produce. See my forthcoming (2019) article for an elaboration of this data structure. A test site that uses Beal for navigating manuscripts can be viewed here:

<http://drc.usask.ca/projects/jon/donne/index.php?table=manuscript>.

A key function of the proposed website will be key-word searching, the affordance most desired by our core target users. I am currently working with programmers at the Digital Research Centre at the University of Saskatchewan to build a search interface using the Apache Solr search platform. By the end of summer 2019, I hope to have in place a basic website that will enable users to locate and navigate texts by:

- work
- document (both manuscripts and print editions)
- key-work searches

Future developments, toward Phase Two

In our initial prospectus, I projected future elements for subsequent phases: a full set of images of all the source documents (witnesses); digital surrogates of landmark materials related to Donne; integrated tools for textual analysis; bibliographies of primary and secondary materials; and provision for user contribution and annotation. Another departure from our original Phase One workplan has been a diversion into annotated bibliographies of scholarship on Donne prepared and published by John R. Roberts. One of our society's members, Jesse Sharpe (previously a digital librarian and now a faculty member at Houghton University), has taken on a project to digitize and parse four volumes (three of them print-born) of these biographies for ingestion into database form. These materials will soon be available for prototyping the next phase of the archive.

As noted above, one of the founding principles of the project, at least in its first phase, was that we wanted to model a project that could be built and sustained without major funding. I think we have come able to a long way following this principle, but by the end of Phase One, we will probably need to seek major funding to take the project forward. Application for funding is probably a couple years away. The first phase would have to be complete before we will have any hope of success with a grant application. An important next step will be to approach the Digital Donne project to work out how these two projects might relate in a future phase of development. I hope also, in the next couple of years, to begin wireframing a fuller, richer knowledge environment for the study of Donne, with the prose and poetry transcriptions at the

core of it, and structured using the primary bibliographies we have prepared (Keynes and Beal) and integrating Roberts's bibliographies.

Works cited:

Beal, Peter, and John Lavagnino, eds. *Catalogue of Literary Manuscripts 1450-1700*.
<http://www.celm-ms.org.uk/>.

"About the Project." Donne Prose Wiki. University of Saskatchewan.
<https://wiki.usask.ca/display/DP/About+the+project>

Digital Donne: The Online Variorum. General editor, Gary Stringer, Texas A&M,
<http://digitaldonne.tamu.edu/>.

Donne Prose. wiki. <https://wiki.usask.ca/display/DP>.

John Donne Society Digital Prose Archive. blog. <https://donneprose.blogspot.ca/>.

Keynes, Geoffrey. *A Bibliography of Dr. John Donne*. Cambridge UP, 2013.

Nelson, Brent. "Radiant Donne: A Case for the Digital Archive and the John Donne's Society's Digital Prose Project." *John Donne Journal* 33 (2013): 175-200.

Nelson, Brent. "The Sociable Textual Archive: Laying the Groundwork for Linked Bibliographic Entities." *KULA: Knowledge Creation, Dissemination, and Preservation Studies* (2019). DOI: <https://doi.org/10.5334/kula.12>.

Roberts, John Richard, editor. *John Donne: An Annotated Bibliography of Modern Criticism, 1912-1967*. University of Missouri Press, 1973.

Roberts, John Richard, editor. *John Donne: An Annotated Bibliography of Modern Criticism, 1968-1978*. University of Missouri Press, 1982.

Roberts, John Richard, editor. *John Donne: An Annotated Bibliography of Modern Criticism, 1979-1995*. Duquesne University Press, 2004.

Roberts, John Richard, editor. *John Donne: An Annotated Bibliography of Modern Criticism, 1996-2008. DigitalDonne: The Online Variorum*.
<http://donnevariorum.tamu.edu/html/resources/robertsbib/Roberts1996.pdf>.

ⁱ This report is based, in part, on a presentation a conference sponsored by INKE (Implementing New Knowledge Environments), Beyond Open: Implementing Social Scholarship, 10 January 2018, Victoria, BC.