

“Pacience is an Heigh Vertu”: Managing *The Canterbury Tales Project* Via *Textual Communities*

Kyle Dase, University of Saskatchewan, kyle.dase@usask.ca

Nicole Atkings, University of Saskatchewan, nicole.atkings@usask.ca

Abstract

Any large digital humanities project presents a difficult institutional problem: a small cluster of academics, most likely traditionally trained as independent researchers, can find themselves at the head of a team that closely resembles a small tech startup. At least, this was the experience of *The Canterbury Tales Project, Phase 2*, with upwards of thirty employees transcribing on an environment under ongoing development; programmers working on that environment ; and senior members of the project promoting that environment and transcription of the *Canterbury Tales* to other academics internationally through workshops.

This article is a reflective essay on the second phase of the *Canterbury Tales Project* and the various successes and challenges that unfolded throughout that process. Our focus is how the project both managed the transcription team working locally at the University of Saskatchewan and facilitated transcription workshops abroad. We detail our training process and the transcription workflow as facilitated via the Textual Communities environment. We also examine and evaluate the causes of the project's challenges—often the result of institutional pressures or technological changes—and our reactions to those challenges, emphasizing successful strategies. Finally, we proffer future changes for the project that we believe would have made considerable positive impact if implemented from the outset of phase two and still have potential as helpful resources now. It is our hope that in detailing our process we can help other large DH projects mimic our successes and, perhaps even more importantly, avoid any pitfalls that challenged us.

Keywords

The Canterbury Tales Project; Project Management; Textual Communities; Digital Humanities; Manuscript Studies; Transcription

Introduction

Any large digital humanities project presents a difficult institutional problem: a small cluster of academics, most likely traditionally trained as independent researchers, can find themselves at the head of a team that closely resembles a small tech startup. At least, this was the experience of the Canterbury Tales Project (CTP), Phase 2, with upwards of thirty employees transcribing on an environment under ongoing development; programmers working on that environment; and senior members of the project promoting that environment and transcription of *The Canterbury Tales* to other academics internationally through workshops.

Moreover, the institutional problem mentioned above tends to be exacerbated by a gap in scholarship: while large DH projects often find outlets to document their practices and guidelines, the hierarchies, workflows, and day-to-day operations are generally absent. Even large full text transcription projects such as *Transcribe Bentham* and *Estoria de Espanna*, though their members have published articles on the process and success of crowdsourcing and transcription, tend to focus more on recruiting and retention than training and revision, with a few exceptions (Duxfield 2018, Causer and Wallace 2012, Causer et al. 2012, 2018). This article is a reflective essay on the second phase of the CTP and the various successes and challenges that unfolded throughout that process. It is an attempt to fulfill our larger mandate to be accessible and transparent by outlining our own organization and process (and to encourage others to do the same). Our focus is how the project both managed the transcription team working locally at the University of Saskatchewan and facilitated transcription workshops abroad. We detail our training process and the transcription workflow as facilitated via the Textual Communities environment. We also examine and evaluate the causes of the project's challenges—often the result of institutional pressures or technological changes—and our reactions to those challenges, emphasizing successful strategies. We try to contextualize the experiences on the CTP in relation to other, similar projects throughout. Finally, we proffer future changes for the project that we believe would have made considerable positive impact if implemented from the outset of phase two and still have potential as helpful resources

now. It is our hope that in detailing our process we can help other large DH projects mimic our successes and, perhaps even more importantly, avoid any pitfalls that challenged us.

Part I: Context and Training

Project History

Before discussing the successes and challenges of phase two, however, it is important to contextualize this period of *The Canterbury Tales Project*. In 2010, the project's leaders, Peter Robinson and Barbara Bordalejo, moved the project to Canada when Robinson took up his post as a professor in English Literature at the University of Saskatchewan. Bordalejo and Robinson had plans to develop a new platform to carry the project forward. These plans ultimately resulted in Textual Communities. Textual Communities (TC) is an online textual editing environment developed by Robinson and initially funded through a CFI Leaders Opportunity Fund and a SSHRC Connection Grant for "Social, Digital, Scholarly Editing", both granted in 2012. In 2014, Robinson received funding from the Social Sciences and Humanities Research Council of Canada for the second phase of the CTP and began transferring the project to the TC environment in preparation for a larger effort to complete the transcription process. This required adding more manuscript images to those the project had already collected, uploading all project images to TC, and aligning the lineation of a base transcription of the text with the corresponding images. From the development of TC onwards, all transcription on the project has been facilitated by that editing environment as our team continues in the goal of transcribing and proofing nearly 30,000 pages present in 88 pre-1501 witnesses. The vast scope of the project is an ongoing challenge, but the transcription team has completed over 7000 pages of transcription since phase two began in 2010 (Nelson and Robinson Forthcoming, 11).

The Canterbury Tales in Context with Other Projects

Ours is hardly the first project to deal with and report on the issues that arise when running large transcription teams. Both the *Estoria de Espanna* and *Transcribe Bentham Projects* have detailed and examined their experiences. The former is a project ultimately aimed at transcribing and collating

the 39 manuscript witnesses of the 13th century chronicle by Alfonso X of Castile, while the latter project seeks to transcribe the unpublished manuscripts of Jeremy Bentham. However, our circumstances are slightly different from those projects. For example, the experience and build of our team is different. *Transcribe Bentham* relies upon crowdsourcing for its transcriptions and, although Jeremy Bentham's hand can be difficult to transcribe and changes over time, its transcribers need to learn to identify only a single hand writing in modern English. *Estoria de Espanna's* transcribers face a challenge much more similar to our own—they must identify and transcribe multiple hands of medieval Castilian—but they also use a large body of crowdsourced volunteers to prepare transcriptions at different levels (Duxfield 2015). Whereas *Transcribe Bentham* relies heavily upon crowdsourced labour for all its transcription, *Estoria de Espanna* sees this labour as only an initial step in preparing transcription for paid transcribers:

In short, we are optimistic that with careful and specific strategies in place to work towards its success, crowdsourcing will enhance and facilitate rather than replace our own in-house produced transcriptions (Duxfield 2015, 138).

The *Canterbury Tales Project*, on the other hand, has not utilized crowdsourcing. While we have allowed and encouraged small contingents of students and volunteers to participate in the project, crowdsourcing requires that “anyone can sign up, regardless of prior qualifications or experience” (Duxfield 2015, 138). We feel that the specialized task of transcribing texts that require training in Middle English paleography and encoding those transcriptions in accordance with the demanding conventions developed by the project requires a level of instruction that excludes crowdsourcing. Instead, our project has depended upon a relatively small cohort of paid transcribers and volunteers who are often trained as part of a class or workshop. This is an important difference from our counterparts: Polly Duxfield has written extensively about the experiences of the *Estoria de Espanna Project*, as have Tim Causer, Melissa Terras, et al. for *Transcribe Bentham*, but both tend to focus on the retention and recruitment of crowdsourced labour or the process of transcription and guidelines more than the practice and structure of their work with paid transcribers (Duxfield 2018, 48–52; Causer and Terras, 2018). That said, there are still many corollaries between our own

experience and those described in these articles and we frequently contextualize the CTP in light of these other important projects below.

The Functionality and Features of the Textual Communities Environment

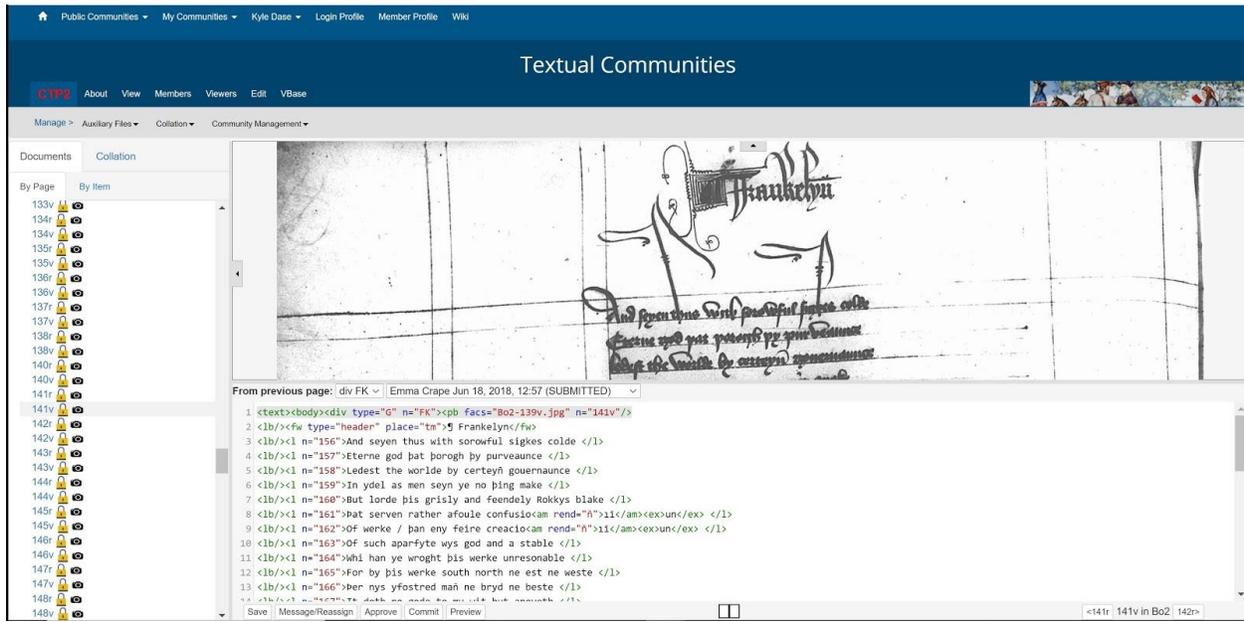


Figure 1: The Supervisor's Interface on the Viewer of the *Textual Communities* Environment.

Before discussing the nature of the project's structure and management in more detail, it is useful to describe the editing environment that facilitated our work, Textual Communities. Bordalejo and Robinson's decision not to develop a transcription environment dependent on existing technology (e.g. the Wikimedia software used by *Transcribe Bentham*) allowed the CTP to create an editing environment that serves its specific needs, with the tradeoff that it would have to develop features freely available to other projects developed on preexisting software (e.g. *Transcribe Bentham*'s "Transcription Desk" and "Benthamometer").

Textual Communities has existed in two versions. The first iteration of TC was built using an open-source platform called LifeRay, which facilitated many features no longer available in the present version such as the chat, wiki, and bulletin board features. However, the project stopped using LifeRay's services for several reasons. Firstly, the project became aware of serious security

flaws in LifeRay, leading to considerable amounts of time having to be spent identifying and remedying breaches. Secondly, we encountered problems with the login and user authentication procedures which appeared insoluble in the LifeRay environment. These difficulties were exacerbated by the withdrawal of support by Google for the fundamental software support underlying LifeRay, when it cancelled its "Google +" social media effort (compare the similar effect on the development of the CantApp when Adobe withdrew support of PhoneBuild, [North et al., Forthcoming]). Simultaneously, the project came to realize that the relational database backend to the first version of TC could not provide the performance required for a live editing environment, and its developers (Robinson and the programmer Xiaohan Zhang) decided to move to a JSON document database backend.

Development on the new TC began in 2014. This was somewhat akin to rebuilding an aeroplane in flight, and it was not until 2018 that the new system was ready, and all existing transcriptions moved to the new system. This change came with some costs. Because the new TC environment focussed on the core tasks of the project (transcription, collation, analysis and publication) it did not attempt to replicate the social media tools built into LifeRay. These tools were either replicated outside TC (thus the Wiki and Chat) or dropped. Two casualties were the bulletin board and ability to compare transcriptions in the viewer page (discussed in more detail below).

The ongoing development of the Textual Communities environment gave our transcription team the opportunity to engage and experiment with these various features. The loss of certain features, such as the chat function, were negligible to the CTP's progress and day-to-day operations because third party software (e.g. *Facebook Messenger*, *Discord*, *Slack*, etc.) provide equal (and sometimes better) functionality at no cost. Likewise, there is little difference now that the TC and CTP wikis, rather than being features of the editing environment itself, are now housed as spaces on the University of Saskatchewan's Wiki Service ("Canterbury Tales Project 2 Home" 2018). In both these instances, transcribers still retain access to the tools that Textual Communities no longer provides directly: transcribers still communicate with their supervisors and one another using chat systems and the entirety of the wiki is freely accessible in its new location through a link within the TC interface.

One feature that showed initial promise but that we have unfortunately lost in TC's second version is the bulletin board. This feature was essentially a forum into which transcribers could upload images of and comments on particularly difficult transcription problems for discussion and resolution among the project community. The thread of discussion then became not only a reference for transcribers who encountered similar problems but a crystallized example of how our project considered and worked through unanticipated questions of our transcription practice. Moreover, the discussion threads were an accessible way for members of the transcription team to learn about our transcription principles in a participatory way and clarify their own understanding of the project. Its role as a hub for the community made the bulletin board a feature that benefitted new and senior transcribers alike.

Another lost feature with clear advantages gave users the ability to compare transcriptions. A user could simply choose two transcriptions in the viewer that would appear side by side and the system would highlight sections where the transcriptions were different from one another. The compare feature was particularly useful for troubleshooting errors in the XML tagging that prevented the transcription from rendering properly. Missing tags become obvious when highlighted in juxtaposition to the most recent functional version of the transcription.

An important feature present in both iterations of Textual Communities is a versioning system that records the activities of its users. Every action a user takes in the system (e.g. saving a change, submitting a transcription, approving a transcription, etc.) is recorded and attributed within TC. The CTP can use this system to be more transparent, held more accountable, and to give credit to individuals for the work they put into the project. Thus, the project publications built on the TC system will give the names of every transcriber who worked on any page of transcription. This can be seen in Thomas Farrell's forthcoming edition of the *Tales of the Reeve and Cook*. The project leader can then also look at important metadata about the amount of time spent on each transcription, the number of pages each person has transcribed, and how the project is progressing overall.

Finally, the most important feature of Textual Communities for the purposes of our transcription team and this paper has been in place since April of 2018 and remains an integral part of the CTP's transcription workflow. The supervisor's interface (and the backend it gives users access to) allows Textual Communities to facilitate and streamline the workflow and exchange between supervisors and transcribers on the transcription team without resorting to external emails (Figure 1). Whereas transcribers have only the options to "Save", "Submit", or "Preview" their transcriptions, supervisors can additionally "Commit", "Message/Reassign", and "Approve" these transcriptions. Under this feature, when transcribers submit their work, an automated email notifies their supervisors of the submission. The supervisors can then review that work and choose to approve the transcription if there are no errors or reassign the submission with a message that details the errors still present in the transcription. The system then sends that message to the original transcribers in an email that both notifies them that their work has been reassigned and what they need to do for that transcription to receive approval. Supervisors also have the ability to make small changes themselves if they deem the errors too minor to warrant sending the page back to the original transcriber. When a supervisor approves a transcription, Textual Communities forwards the page to the project leader in much the same way as it does with a transcriber's work. The project leader then reviews the page one last time before committing it to the system as ready for collation.

This workflow is far better than the system of emails that existed before it. In the earlier system, each transcriber would have to write and send an email from his or her own account to a supervisor and the messages did not have the uniformity they now do. A supervisor would likewise have to compose many emails. Worse, transcriber emails from multiple accounts could sometimes get lost. The new system sends all correspondence through a single, easily searchable email address. Beyond the benefits of an automated system that saves time for both transcriber and supervisor, the new interface is more organized as well. Each folio page and its place in the transcription workflow is recorded and displayed for each user in that user's member profile (Figure 2). One can even enter back into the viewer to transcribe or review a specific transcription by selecting its link in the list

on the member profile page. While a straightforward addition, this simple interface and the programming behind it have been the greatest improvement to the project's efficiency and workflow for the transcription team.

Name	Role	Joined	Assigned	In Progress	Submitted	Approved	Committed
CTP2	LEADER	Sun Jan 11 2015	(63) • Ad3:145r • Ad3:244v • Ad3:256r • Ad3:264r • Ar:183v • Bw:84v • Bw:85r • Bw:85v • Bw:86r • Bw:86v • Bw:87r • Bw:87v • Bw:88r • Bw:88v • Bw:89r • Bw:89v • Bw:90r	142 • Ad3:26r • Ad3:26v • Ad3:27r • Ad3:245v • Ad3:246v • Ad3:247v • Ad3:248v • Ad3:249v • Ar:181v • Ar:182r • Ar:182v • Ar:183r • Bo1:211r • Bw:9v • Bw:10r • Bw:13v • Bw:14v	746 • Ad1:27r • Ad1:29r • Ad1:31v • Ad1:36r • Ad1:108r • Ad1:108v • Ad1:109r • Ad1:110r • Ad1:110v • Ad1:286r • Ad1:287r • Ad1:287v • Ad1:288r • Ad1:288v • Ad1:289r • Ad1:289v	579 • Ad1:19r • Ad1:19v • Ad1:20v • Ad1:21r • Ad1:30r • Ad1:32r • Ad1:32v • Ad1:33r • Ad1:33v • Ad1:35r • Ad1:35v • Ad1:37r • Ad1:37v • Ad1:38r • Ad1:38v • Ad1:40r • Ad1:40v • Ad1:41r	894 • Ad1:13r • Ad1:13v • Ad1:14r • Ad1:14v • Ad1:17v • Ad1:18r • Ad3:3r • Ad3:3v • Ad3:6r • Ad3:6v • Ad3:7r • Ad3:7v • Ad3:8r • Ad3:8v • Ad3:9r • Ad3:9v • Ad3:10r

Figure 2. Member Profile Page for a Transcription Supervisor in *Textual Communities*.

While some of these features were more successful than others, reflecting on each helps us to understand just how the project and its members function. In particular, it brings into focus our nature as a community that requires opportunity for intellectual exchange and discourse as part of our work, a need that many of our best features have aimed to meet and make more efficient. This is not something one would necessarily assume of work that can be done in isolation like transcription.

Training the Transcription Team for CTP Phase 2

One of the interesting challenges the project faced was how to get transcribers of various different specializations and backgrounds up to speed about the overall purpose of the CTP and the role transcription plays in the larger project. Transcribers were sometimes senior undergraduates who had taken a course in digital humanities, MFA in writing students, MAs in English, or doctoral candidates who study medieval literature. The varying levels of background required us to develop

training materials and practices that ensured every transcriber knew enough to be a productive member of the team without overextending our senior project members. We had to make sure not to presume our transcribers had greater knowledge than they really did while expediting their training so that they could contribute to the project. Training tools such as sample pages, wikis, and quizzes supplemented the face-to-face training of supervisors as a means to efficiently bring our team up to speed.

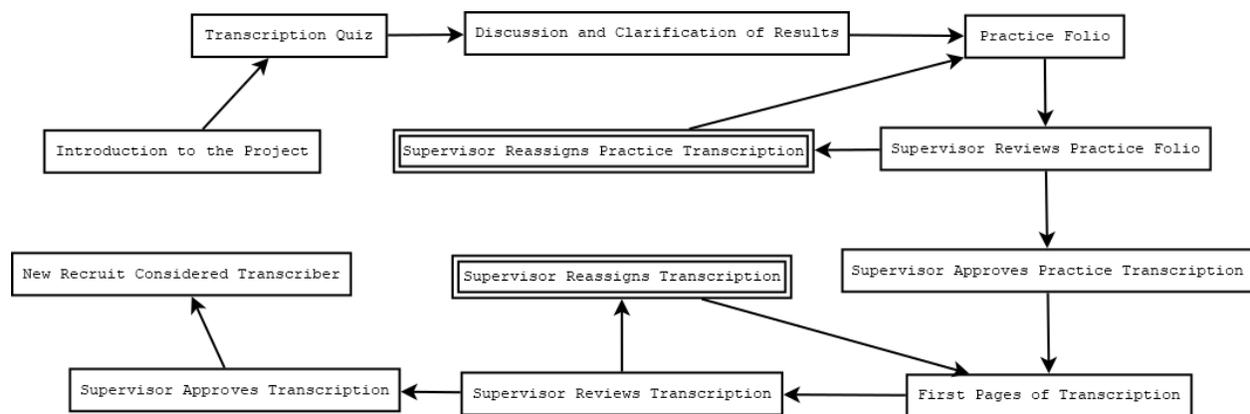


Figure 3. Current Transcriber Training Workflow.

Transcriber training begins with an in-person meeting where a project leader or transcription supervisor provides a general introduction to the project. Locally, this usually entailed a one-on-one meeting with a new employee (though this interaction could sometimes take place during the group meetings described below). However, a large enough group of new recruits could result in a small workshop similar to those delivered abroad (see section below). Some of the most important training for our transcribers was quite general compared to the niche details of transcription practices one might expect us to prioritize. For instance, it became important to make sure each transcriber had a proper understanding of our ultimate aim of collating electronic transcriptions as it illustrated the different degrees of significance for certain tasks. Otherwise, as occurred in one instance, a transcriber might sometimes agonize over representing minutiae that would be regularized during collation but miss an entire word in the same line instead. Proper context of the larger task at hand is paramount for a transcriber's success.

Generally, we explained the basics of the project to each new transcriber: Chaucer's *Canterbury Tales* exist in various manuscript witnesses from the Middle Ages and most modern editions work from a select few copy texts. We instead seek to collate that manuscript tradition in its entirety. From this explanation, a brief discussion of the collation process and the project's earlier work with phylogenetic software was usually enough to give transcribers an idea of why and *how* their work matters (Barbrook et al. 1998). In other words, they were ready to practice transcribing. At this point, the supervisor will display a sample transcription and explain the basic practice of transcription, the meanings of TEI and XML, how our transcription works, and introduce the new transcriber to our transcription guidelines. Once the new transcriber has read over the "Quick Start Guidelines" and "Full Transcription Guidelines", he or she is ready to take the transcription quizzes.

One of the first tools we implemented for transcriber training was an online transcription quiz (still available at <https://wiki.usask.ca/display/CTP2/Do+the+quizzes>). Once transcribers had read over the guidelines and had a chance to consider how they might implement those guidelines in practice, the quizzes—one introductory and one more advanced—provided a low-risk context in which we could make sure new transcribers understood the most rudimentary aspects of transcription. The questions ensured comprehension of fundamental practices such as that an abbreviation tag (<am>) should have a corresponding expansion tag (<ex>) or that we record "ff" at the start of a word as a capital "F". Rather than having to wait for the transcriber to encounter a particular problem out in the wild or having them transcribe another entire practice page when all we wanted was to test their knowledge of a single instance, the quiz allowed us to cover many issues in a format that generally takes the transcriber only a short time and delivers the correct answer. We found the quizzes particularly useful in environments where multiple new recruits were taking the quiz and an authority on the project was available to provide clarification for the transcriber's misconceptions and generate discussion among transcribers about where their intuition may have led them astray.

After a new recruit has finished the quizzes and demonstrates an understanding of their content in the subsequent discussion, we then assign him or her a practice manuscript. This practice manuscript is part of a separate community in TC called “CTP Training”. This community contains a single “manuscript” called Sample Page that is merely 399 iterations of the same practice folio (Hg 223r) and we assign an iteration for each new transcriber to practice on. This system has several advantages: first, we can choose a page that the project leaders have considered in advance. They can select an example that has a diverse range of common and challenging transcription problems that require consultation of the guidelines. We can also be sure that the folio is an appropriate challenge level for a beginner. Our current practice folio requires a new transcriber to encode an ornate capital, a header, some of the most common abbreviations found in the manuscripts, and even a straightforward use of the <app> element (see Bitner and Dase Forthcoming, 13–14). Moreover, this strategy allows us to train new employees more quickly, as we have an answer key to their first transcription and can point out any serious issues before they begin independently transcribing material.

Once a transcriber has taken the transcription quiz and perfected the practice folio, we begin assigning them manuscript pages. At first we assign a small batch of four or five folios because, just as with the practice folio, we want to catch any of the new transcriber’s errors before a large number of folios is transcribed. Our supervisors try to review and give feedback as quickly as possible on these first transcriptions since we only begin assigning new pages to a transcriber after his or her supervisor is satisfied with and has approved those initial pages. In some rare cases, if a transcriber is having an especially difficult time, he or she may repeat this step with another small batch of transcriptions. Generally, however, there is a smooth transition from this step to our regular process of transcription and revision. A supervisor tries to keep in mind a transcriber’s level of experience and review transcriptions in a timely manner. Ideally, a paid transcriber will eventually become well trained enough that he or she can either become a supervisor or submit his or her transcriptions directly to the project leader.

While this is our current training process, it was not our first iteration nor will it necessarily remain unchanged. For instance, we experimented with different training periods for transcribers throughout the project, beginning with more attentive transcription supervision for our earlier employees and reducing that training period as we took on more transcribers. The first new transcriber, working under the supervision of a doctoral student, submitted nearly 100 pages for revision before submitting his work directly to the project. Part of the reason this transcriber was subject to such scrutiny was that the full quiz and practice pages were not yet in place and his supervisor wanted to be sure his work was accurate. With no official method in place, ongoing supervision seemed the most plausible route. Still, this was likely overcautious and subsequent transcribers had somewhere between a third and half as many pages checked with the meticulousness that initial training requires, pending their performance on transcription they had already done.

Workshops Abroad

In addition to managing a local team of transcribers along with volunteers from abroad, our project has also delivered workshops at other academic institutions for students and faculty interested in working on the CTP. In our most recent workshop, delivered at Duke University in September 2018, we taught approximately fifteen senior undergraduate and graduate students in collaboration with their Center for Medieval and Renaissance Studies. The training provided in the workshops is a condensed version of the training we give all of our transcribers which is then followed up through online correspondence.

First, the participants learned the basics of our transcription guidelines and our team led them through a tour of the features available on the TC environment. Once participants became familiar with the interface and the guidelines, we ran through some examples of challenges they might face during transcription while fielding any questions they had. At this point, we gave the participants the same quizzes we normally use to train our transcribers as mentioned above (see section 4). After the participants receive their results, we discuss the correct answers and the specific transcription challenges the questions are meant to raise before turning to the practice

transcription (see section 4). At this stage of the workshop, participants often begin to work in collaboration with one another, helping one another with the transcription. Once again, we go over this transcription practice as a group and use it as an opportunity to clarify any ambiguities that participants might still have about the guidelines. Finally, we provide the participants with a few uncurated pages to transcribe while we are present to round out the workshop and continue to field questions and review their transcriptions through a combination of e-mail and the TC environment.

The structure of this workshop is designed to optimize the face-to-face time between workshop coordinators and new transcribers in order to properly train the transcribers and weed out the most common and problematic errors. We generally encourage these cohorts of satellite transcribers to meet with one another once a week in the same way our own group does to support one another in their transcriptions and build a sense of community. We find that maintaining face-to-face communication, just as with our own transcribers, makes them more likely to continue transcribing consistently.

We should also be clear that the CTP project does not envision the training of its transcribers as ending with this initial training process. The project offers weekly meetings in which transcribers can bring difficult challenges they encounter in their work for a group resolution. These challenges often clarify our own practices to meeting attendees who observe and even participate in the discussion though, just as Polly Duxfield explains of the process on the *Estoria de Espanna Project*, final decisions on points of contention rest with the project's leaders (2018, 56-58). Project leaders also sometimes deliver short lectures at these meetings to give the CTP context or better illustrate the principles behind a particularly nuanced transcription practice such as the <app> tag. As discussed above, transcribers' experiences with the bulletin board constituted a form of participatory learning in a similar way to the discussions in group meetings. Even the collaborative attitude of the project's members sees our transcribers frequently transcribing in small groups and partnerships, teaching one another and reinforcing what they have already learned (Nelson and Robinson Forthcoming, 14-15). The first steps towards becoming a

full transcriber are an initiation into a larger body of team members who are continuously learning from one another through the collaborative transcription process.

Part II: Challenges and Growth

The following sections concern the growth of our transcription team, the challenges of our specific circumstances, and suggestions for how we might have handled certain challenges differently in hindsight as well as for improvements or features that might still benefit the project. We look at the unique situations of student employees and the fluctuation in their available working hours as well as the institutional circumstances that put pressure on our own project to work at a larger scale and faster rate of transcription. We also raise the possibility that there are gaps in our current support structures that the project could fill for a more satisfactory communication between transcribers and managers. Our goal with these suggestions and observations is not merely to criticize our own project (though we do believe we should look at the CTP with a critical lens) but to help us frame those qualities that make us an effective collaborative transcription team.

Growth of the CTP

The project's initial growth was measured and relatively small. At first, just a half dozen transcribers were trained by the project leaders themselves and learned about the greater context of the project in weekly meetings while transcribing. From there, we gradually hired more transcribers who were trained by the senior transcribers who would eventually become project supervisors (the authors of this paper). At this stage, we were able to oversee each trainee's transcription in a timely fashion. This was not only to the benefit of the trainee: every folio of transcription a trainee encoded in the time between their submitting pages and our review of his or her work would carry early misunderstandings or errors and require time and the resources to go back and correct. This cost is negligible when the number of new employees is small and delays can be kept to a minimum, but can prove an issue when the number of new transcribers greatly outweighs that of transcription supervisors.

At this point in the project, the term limit on the CTP's grant from the Social Sciences and Humanities Research Council of Canada was approaching at a quicker rate than our transcription cohort was using the allotted funds. The project began hiring more transcribers from a wider pool of students in an effort to achieve as much transcription as possible before the project's formal end in May 2019.

The CTP tended to hire graduate students as transcribers when available, however, when we could not secure graduate students for transcription work, we turned to senior undergraduate students with some manuscript experience. For instance, the project hired senior students in the University of Saskatchewan's Classical, Medieval, and Renaissance Studies (CMRS) program. Many of these students had taken some combination of the courses "Exploring Medieval and Early Modern Manuscripts", "Advanced Manuscript Studies", and English literature courses that focus on Chaucer. In some cases, the work of the manuscript courses directly dealt with materials and content of the CTP (Nelson and Robinson Forthcoming, 8). These courses provided this group of transcribers with valuable experience and understanding of transcription as well as some background in manuscript studies and transcription practices with texts from the Middle Ages in addition to a familiarity with Chaucer's work. However, our needs for even more transcribers meant we also hired students who had not received this same training around the same time. While it was necessary to cast a wider net to facilitate a larger workforce, this required condensed training that provided context for the project, rudimentary training in manuscript studies, and a passing familiarity with Chaucer and Middle English beyond what one experiences in an introductory literature survey course. All this, of course, was in addition to the transcription training every transcriber received.

This influx of new transcribers was a definite benefit to the project and resulted in an overall greater rate of transcription but presented some challenges for managing the day-to-day operations. Hiring transcribers with little specialized knowledge required more time spent training them and the pages of transcription under review by the supervisors required closer inspection to protect the overall quality of the transcriptions being produced. These supervisory tasks, when

coupled with the fact that there were more total transcribers requiring supervision, overloaded our supervising team for a time. This experience is by no means unique: the *Transcribe Bentham* team describes a similar instance where a *New York Times* article inspired a surge in new volunteers over the winter holidays when several members of the team were on leave, resulting in a similar problem (Causer et al. 2012, 130). Whereas the consequence for that project was a low retention rate for new volunteers and a backlog that lasted around 10 days, our own was a considerable transcription backlog requiring review and resubmission. The errors made were often fairly insignificant, but the inability to address these immediately meant that they continued to appear in ongoing transcription. This necessitated extra work for the pages' original transcribers when the supervisors were forced to reassign numerous pages on account of a large number of small but oft-repeated errors. As described in more detail below, if we had anticipated this eventuality, we might have put more senior transcribers in place as supervisors before beginning our hiring surge.

This funding issue was further complicated by the ephemeral nature of student employment. Students present a unique challenge when trying to anticipate future progress and regulating the workflow of transcription and revision since their availability can be limited. On its own, this is a relatively minor issue. However, perhaps a more difficult problem altogether is that student schedules fluctuate greatly throughout the term and a student that can contribute only an hour a week in mid-November might work a dozen hours in mid-December. Anticipating the incoming workload for our project leaders and transcription supervisors was frequently a challenge when the project had more transcribers, especially given that supervisor's schedules could be subject to the same kinds of fluctuations. Unlike a transcriber, who can work as much or as little as their availability allows, supervisors need to respond promptly to submissions.

Student employees also frequently leave. Many members of the cohort of phase two transcribers were senior undergraduates near the end of their programs and masters students who had a maximum of two years of study and most often one. As a result, some transcribers would have only just begun turning in their best work when they completed their studies or moved on from the CTP to full time jobs elsewhere. Obviously, this is not the students' fault: the nature of

these work patterns is often the result of the precariousness of student positions combined with the demands on students' time both for their studies and, in some cases, the multiple forms of employment tuition forces them to take. That said, the project did eventually establish a core group of transcribers, several of them PhDs, who participated in the project for well over a year or more, but the circumstances of student employment that affected many members of the team proved to be a consistent issue for the project as a whole.

What We Might Have Done Differently

Had we had the hindsight on the project that we do now, there are several changes we might have made to improve both the quality and efficiency of our transcription process and team. Unlike the *Bentham* and *Estoria* projects, the CTP, as originally funded, did not anticipate use of either "crowdsourcing" nor a cohort of paid student transcribers. Instead, the project expected to use graduate students engaged in doctoral and masters studies related to the project to do the bulk of transcription, as had been the case before the project moved to Canada. When the project was not able to recruit these graduate students, it changed course to the employment of paid student transcribers. Many of the problems here identified are the consequence of this shift of direction.

First, it would have been beneficial to have trained transcribers who could stay attached to the project with the intent to promote them to supervising and revising the transcriptions of others. As Polly Duxfield notes in "Transcribing the *Estoria de Espanna Using Crowdsourcing*", it takes time and resources to initially train a transcriber to the point that he or she can expedite transcription for those overseeing his or her work:

It is important to remember, however, that in order for volunteers to reach this level a significant amount of time will have already been invested in developing training materials and in mentoring the transcriber. It is only after this time investment that it becomes quicker to check volunteer-transcribed folios than it does to transcribe them in-house (2015, 138).

Although our project took on paid transcribers, Duxfield's point remains relevant. Especially in the earlier stages, it would have been prudent to have sought transcribers who could commit to longer terms of employment. We could have communicated the need for these kinds of roles in the project and gauged interest from potential candidates. While many facets of a supervisor's role are the same—one's primary role is still looking over and correcting a transcription—there are certain constraints that make the job less amenable to certain candidates. For instance, transcribers can, for the most part, work according to their own schedule, an ideal job feature for most students. However, supervising transcription requires the supervisor to work in response to the transcriber, revising and delivering feedback on transcription in a timely manner to ensure both that the transcriber has a chance to fix any errors and that those errors do not carry over into their future transcriptions. This requirement for a quick response time, as well as the added step of communicating with others working on the project made some members of the project who either already had another job with a rigid schedule or needed a more flexible schedule to prioritize other academic work prefer to continue working as transcribers rather than supervisors. As some transcribers preferred to do their own transcription without having to worry about supervising the work of others, two different employment tracks could have been implemented from the beginning: one where participants could focus on turning over a large amount of personal transcription that the project leaders could check periodically, and another where participants were either matched in pairs or small groups to review each other's work under the guidance of the project supervisors. Had we communicated our needs more clearly from the outset of phase two, we may have been able to secure dedicated transcribers interested in becoming supervisors who felt secure in the promise of ongoing work and contributed a level of stability to the project, more reasonably distributing revision duties during particularly busy periods of transcription.

A more stable employee hierarchy and infrastructure would also have allowed for more consistent budgeting. As we already mentioned, the project was trying to achieve as much transcription as possible by the project's formal ending. This instigated a hiring surge to ensure that we could be as efficient as possible before the deadline. However, this rapid accumulation of new transcribers on the project brought about challenges of its own, as discussed above. While

many of these variables were out of our control, the project could have benefitted from work and budgeting projections such as a Gantt chart or simply budgeting the project backwards from our project deadline based on employees' average monthly hours.

The project also could have benefitted from a more explicitly applied pedagogical structure within our weekly meetings. For instance, the project leaders encouraged transcribers to bring questions and challenging transcription issues to transcription meetings so that all transcribers could benefit from resolving the issue. As stated above, this is a valuable practice for the transcription team that should remain. However, had we requested that transcribers submit their issues ahead of time, supervisors could have screened submissions and gone over issues that were more valuable for group discussion with project leaders and left other issues for email correspondence or even something similar to office hours for supervisors. Examining issues in advance would have allowed us to have covered more issues in less time and to spend the rest of our group meetings educating the transcription team on other aspects of the project or actively transcribing in the same physical space. It would also have been valuable to organize and prioritize (especially for those with schedules that did not allow for regular attendance) a monthly meeting where project leaders might discuss such matters as overall progress for the project, as well as the dissemination of information about the manuscripts currently undergoing transcription, and any important changes to transcription practice or issues the transcription team might consistently have problems with. These changes would have made our meetings more efficient in general, facilitating communication between transcribers and project managers and leaders.

What We Might Do Better

While there are certain challenges and experiences that we might have handled better in hindsight, we also need to keep in mind the kinds of features that could make our own work more efficient and effective. The suggestions below are meant both to illustrate features that would have ameliorated those past experiences and challenges as well as being productive for the project at its present phase. The problems they flag and identify are as important to this article as the features themselves as a means of isolating and analysing what makes for a productive and efficient project.

One significant way to improve the project would be an inclusion of a common errors section in the project wiki addressing frequent anticipatable issues that arise for new transcribers. While the quickstart transcription guide can serve this function with straightforward transcription issues, a common errors guide could lead transcribers through more complicated issues that require more nuanced explanation and troubleshooting but are not appropriate as part of the quickstart guidelines. While the bulletin board, when it existed, did satisfy this need to some extent, a common errors section would also have unified our response to recurrent issues which the collaborative nature of the bulletin board would sometimes confuse or undermine.

The creation of a wiki that features common errors such as this one could serve as both prerequisite reading in training and workshops in addition to a first stop for transcribers who encounter a problem they are unsure how to solve. We could even recover some of the more prominent threads from the now defunct bulletin board mentioned above and reframe the solutions there as entries for the new wiki.

We also see a need to develop resources or modules that help uninitiated and unspecialized transcribers to better understand the nature of our work. The transcription team might benefit from a set of resources and modules that both summarize scholarly debates on practices in textual editing and transcription in order to contextualize our own practice, as well as more introductory materials on Chaucer and his work. In its simplest form this might constitute adding another prerequisite document to the CTP wiki. However, there are full text transcription projects that have created more engaging video content instead. For instance, the *Estoria de Espanna Project* has found success in creating YouTube videos that provide the necessary context for their users, with subjects ranging from “the manuscript that we are going to transcribe and the main features of its materiality and writing” to “advanced features of the transcription tool” (“Module 1”, “Module 4”). In describing the rationale for the research modules of the *Estoria de Espanna Project*, Polly Duxfield explains another benefit to providing such materials in an online format:

it is much easier and less time-consuming in the first instance for the staff-member to direct the volunteer to a certain module of the course than to explain a particular tagging issue on a one-to-one basis (Duxfield 2015, 144).

Admittedly, such videos as the *Estoria de Espanna Project* provides require a considerable amount of work to be produced well and often have a greater payoff for large crowdsourced projects. However, we believe something similar could be a good fit for our project. Especially for new recruits without much background in textual editing or medieval literature, a small series or even a single introductory video on *The Canterbury Tales* as a work and the CTP's broader aims and methods would have been helpful. Moreover, this could save project leaders and supervisors from repeating the same information in person and prioritize face-to-face interactions for further discussion of the video's content and engagement with the practice of transcription. This would have been an important resource during our hiring surge (see p. 11) and could have mitigated some of the mistakes present in the backlog generated from that experience.

Finally, we believe that the transcription team would have benefitted from a greater knowledge of the project's ongoing progress. When working on such a large endeavour that has been ongoing for over a quarter century, it can be difficult to understand the scope of the work one is participating in. At its worst, transcribers can begin to perceive each page of transcription as an isolated task. Earlier phases of the project focussed on individual sections of *The Tales* at a time and completion was often marked with the release of an edition. We could have new recruits engage with the editions already published in collaboration with the CTP. A quick glance through these editions can better illustrate our goals in more concrete terms. We also posit that, in addition to acknowledging and supporting individual rates of completion (something that earlier versions of Textual Communities did explore through rudimentary rankings and gamification in the bulletin board), the project's leaders should make an effort to celebrate and acknowledge certain transcription milestones to give the transcription team a sense of accomplishment over time. Crowdsourcing projects such as *Transcribe Bentham* offer some precedent for this. That project features a "Benthamometer" on its website that displays the ongoing progress for the entire project

on its website as well as progress for each specific box of Bentham's papers (Duxfield 2015, 135; "Benthamometer"). One could easily imagine a feature in TC that performed a similar function for *The Canterbury Tales* as a whole and perhaps even the progress of individual manuscripts or tales. In fact, TC is already recording much of that data, as Robinson has shown elsewhere (Nelson and Robinson Forthcoming, 11). It may even be in the best interest of morale and camaraderie to organize transcription assignments around individual sections, tales, or even subsections of particularly long tales. Indeed, we have even taken to this practice when coordinating workshops with other institutions (e.g. the team at Duke has been transcribing *The Wife of Bath's Tale*). Assignments could be for the entire project or one could even envision assigning small transcription teams specific short-term goals to give transcribers a better sense of accomplishment that indicates why their work matters. Though we realize that such organization could potentially result in possessiveness over intellectual property—a problem Robinson has been happy to avoid since arriving at his post at the University of Saskatchewan (Nelson and Robinson Forthcoming, 14–16)—we believe organizing transcription in this way could actually improve collegiality in our current environment and the transcription team's sense of accomplishment.

The suggestions outlined above are features that would not only improve the quality of life for the project's members, they also speak to important fundamentals of communication between project organizers and transcribers. Causer et al. posit that volunteers may "feel undervalued, or exploited" if those supervising transcription do not provide meaningful feedback and quality control (2012, 130). This statement is equally true for paid transcribers and we would add that a good project must communicate an appreciation of the progress its transcribers contribute to or run that same risk of alienation. Likewise, we must take every opportunity to effectively communicate project expectations to transcribers. The CTP already encourages its transcribers to be heard through its frequent in-person meetings and the availability of its project leaders and supervisors for correspondence. The progress meters, common errors guide, and instruction modules would therefore help us to assure transcribers that their efforts are appreciated and would communicate our own expectations for their work more clearly.

Conclusion

Traditionally trained academics need all the help they can get coordinating and facilitating the work of large teams. In hindsight, the most important lessons for project management throughout our experience on *The Canterbury Tales Project* have revolved around how we conduct ourselves as a community and negotiate the dynamics between community members. The Textual Communities environment, as its name implies, exists to help us mediate these tasks. Indeed, the best of its features discussed in this paper facilitate the relationships between transcribers, whether that means the bulletin board's ability to facilitate group exchange or the way that the TC's interface and backend help transcribers and supervisors communicate in a cleaner, more effective manner.

An environment like Textual Communities can be a powerful tool to help facilitate a large project, but problems still inevitably arise. We have learned from these challenges and hope that others can learn from our own experience. Often, these issues are the effects of pressures from institutional structures such as funding agencies or universities. Student schedules can create unique workflow issues and projects can have seemingly arbitrary funding deadlines stipulating resources. And, though one might hope that certain problems be resolved at an institutional level, anticipating these issues as project coordinators, even being aware of their existence, make them much more manageable in the meantime. Even with the growing pains the project has faced over the years, working on the CTP has been a rewarding experience. By continuing to adapt our procedures and practices, we hope to continue our progress on the more than 7000 pages of transcription completed in the last ten years, looking forward to the day when we can celebrate the complete transcription of the 88 pre-1501 witnesses of *The Canterbury Tales*.

Acknowledgements

We'd like to thank the leaders of the *Canterbury Tales Project*, Barbara Bordalejo and Peter Robinson, for their patience and support as we wrote this article, answering numerous questions and prompting many discussions that led to a more in-depth presentation of our work on the project. We would also be remiss not to acknowledge all the many transcribers on our CTP team whose *patience* in dealing with us as we learned to be better project supervisors can only be described as *an heigh vertu*. Moreover, we thank and acknowledge the funding of the Social Sciences and Humanities Research Council of Canada that made our employment possible.

Author Contributions

- **Conceptualization:** kd, na;
- **Writing – Original Draft Preparation:** kd, na;
- **Writing – Review & Editing:** kd, na;

Works Cited

- Barbrook, Adrian C., Christopher J. Howe, Norman Blake, and Peter Robinson. 1998. "The Phylogeny of The Canterbury Tales." *Nature* 394 (6696): 839.
<https://doi.org/10.1038/29667>.
- "Benthamometer." 2020. Transcribe Bentham: A Participatory Initiative. June 2020.
<http://transcribe-bentham.ucl.ac.uk/td/Benthamometer>.
- Bitner, Kendall, and Kyle Dase. Submitted 2020. "A Macron Signifying Nothing: Revisiting *The Canterbury Tales Project* Transcription Guidelines." *Digital Medievalist*.
- Causer, Tim, and Valerie Wallace. 2012 "Building A Volunteer Community: Results and Findings from Transcribe Bentham." *Digital Humanities Quarterly* 6(2).
<http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>.
- Causer, Tim, Justin Tonra, and Valerie Wallace. 2012. "Transcription maximized; expense minimized? Crowdsourcing and editing 'The Collected Works of Jeremy Bentham'." *Literary and Linguistic Computing* 27 (2): 119–137. <http://doi.org/10.1093/lc/fqs004>.
- Causer, Tim, Kris Grint, Anna-Maria Sichani, and Melissa Terras. 2018. "Making Such Bargain': Transcribe Bentham and the Quality and Cost-Effectiveness of Crowdsourced Transcription." *Digital Scholarship in the Humanities* 33(3): 467–87.
<https://doi.org/10.1093/lc/fqx064>.
- Duxfield, Polly. 2018. "The Practicalities of Collaboratively Digitally Editing Medieval Prose: *The Estoria de Espanna* Digital Project as a Case Study." *Digital Philology: A Journal of Medieval Cultures* 7(1): 46–64. <https://doi.org/10.1353/dph.2018.0003>.
- . 2015. "Transcribing the 'Estoria de Espanna' Using Crowdsourcing: Strategies and Aspirations." *Magnificat Cultura i Literatura Medievals* 2: 129–48.
<https://doi.org/10.7203/MCLM.2.4977>.

Farrell, Tom. Forthcoming. *The Tales of the Reeve and The Cook*. Scholarly Digital Editions.

<http://www.inklesseditions.com/TCP/Open/RE/>.

Nelson, Brent and Peter Robinson. Forthcoming. "Undergraduate Curricular Contexts for Research in Textual Studies" *New Technologies in Medieval and Renaissance Studies*.

North, Richard, Peter Robinson, Barbara Bordalejo, and Lina Gibbins. Submitted 2020. "Turning the General Prologue into an App." *Digital Medievalist*.

Robinson, Peter. "Canterbury Tales Project 2 Home." Last modified April 10 2018. *Usask Wiki*.

<https://wiki.usask.ca/display/CTP2/Canterbury+Tales+Project+2+Home>

———. "Do the Quizzes." Last modified September 20, 2018. *Usask Wiki*.

<https://wiki.usask.ca/display/CTP2/Do+the+quizzes>

———. n.d. *Textual Communities*. Accessed June 23, 2020. <http://textualcommunities.org/app>.

———. 2013. "The History of Scholarly Digital Editions, Plc." *Papers of the Bibliographical Society of Canada* 51(1): 83–104. <https://doi.org/10.33137/pbsc.v51i1.20763>

———. "The History of the Canterbury Tales Project." Last modified April 1, 2018. *Usask Wiki*.

<https://wiki.usask.ca/display/CTP2/The+History+of+the+Canterbury+Tales+Project>

Robinson, Peter, and Elizabeth Solopova. 1993. "Guidelines for the Transcription of the Wife of Bath's Prologue." In *The Canterbury Tales Project occasional Papers*, edited by Norman Francis Blake and Peter Max Wilton Robinson, I: 19–52. Oxford: Office for Humanities Communications. <http://wiki.usask.ca/display/CTP2/Original+Transcription+Guidlines>.