Transcribing the *Estoria de Espanna* using crowdsourcing:
Strategies and aspirations

Polly Duxfield

University of Birmingham

PLP474@bham.ac.uk
http://orcid.org/0000-0003-4656-7283

ABSTRACT

This paper examines the specific strategies for recruitment and retention of volunteer transcribers in use in two collaborative transcription projects: *Transcribe Bentham* (University College, London) and the *Estoria de Espanna* Digital Project (University of Birmingham).

The aim of the paper is to review the strategies used by *Transcribe Bentham*, a more mature crowdsourced electronic transcription project, with a view to informing the strategies put into place in the *Estoria* project, which has started transcribing using crowdsourcing more recently. The paper discusses the difficulties faced by crowdsourced electronic transcription projects and how these have been and are being resolved in these two projects.

The difficulties discussed include the complexity of the palaeography involved, the necessity of tagging transcriptions using XML, the requirement to moderate and carry out quality-control of volunteer-produced transcriptions, and the creation of an atmosphere of camaraderie amongst staff-members and crowdsourcers, many of whom have never, and will never meet face-to-face. The findings may be useful for other collaborative electronic transcription projects and will inform and shape the way the *Estoria* project continues to use strategies to raise levels of recruitment and retention of crowdsourced transcribers. An earlier version of this paper was first presented at the 2nd Annual *Estoria de Espanna* Digital Project Colloquium, Magdalen College, University of Oxford, 14-15th November 2014.

Transcriure l'*Estoria de Espanna* emprant proveïment participatiu: estratègies i aspiracions

RESUM

Aquest article examina les estratègies específiques per al reclutament i manteniment de transcriptors voluntaris en dos projectes de transcripció col·laborativa: *Transcribe Bentham* (University College, London) i el projecte digital *Estoria de Espanna* (University of Birmingham).

El propòsit de l'article és considerar les estratègies emprades per *Transcribe Bentham*, un projecte electrònic de transcripció en proveïment participatiu més veterà, per tal d'informar de les estratègies emprades en el projecte *Estoria*, que ha començat a transcriure emprat proveïment participatiu més recentment. L'article comenta les dificultats amb què s'enfronten els projectes de transcripció electrònica per proveïment participatiu, i com han estat i són resolts en aquests dos projectes.

Les dificultats comentades inclouen atenció a la complexitat paleogràfica que això implica, a la necessitat d'etiquetar transcripcions emprant XML, al requeriment de moderar i mantenir un control de qualitat en les transcripcions generades per voluntaris, i a la creació d'un ambient de companyonia entre els membres de l'equip i els proveïdors participatius, molts dels quals no es coneixen en persona, i potser no es coneixeran mai. Les troballes poden ser útils per a d'altres projectes electrònics de transcripció col·laborativa, i informaran i conformaran la manera en què el projecte *Estoria* continue emprant estatègies per a augmentar els nivells de reclutament i de manteniment de proveïdors participatius. La primera versió d'aquest article va ser presentada al 2nd Annual *Estoria de Espanna* Digital Project Colloquium, Magdalen College, University of Oxford, 14-15th November 2014.

PARAULES CLAU

Proveïment participatiu, transcripció, edicions electròniques, transcripció col•laborativa, voluntaris, XML, literatura espanyola medieval, cròniques, Alfons X, *Estoria de Espanna*, hispanisme britànic, humanitats digitals.

TABLE OF CONTENTS

⇒✳⇐

1 Introduction

The *Estoria de Espanna* Digital project ("*Estoria* project" for short) is a four-year Arts and Humanities Research Council (AHRC) funded collaborative project based at the University of Birmingham, UK, led by Dr Aengus Ward.[1] Our principal aim is to produce a fully-collatable digital edition of the medieval chronicle the *Estoria de Espanna*, which will be freely available online under a creative commons licence.[2] The current phase of funding will see the transcription of five of the currently thirty-nine known extant manuscripts of the *Estoria*—namely: E1, E2, Q, Ss and T[3]—which will then be electronically collated to allow scholars to study the similarities and differences between manuscripts to examine issues such as the rewriting of history, linguistic evolution and the nature of medieval textuality. The *Estoria* project transcriptions and collations are carried out using an online collaborative transcription platform called Textual Communities. This initiative is being developed by scholars at the University of Saskatchewan, Canada, and is led by Professor Peter Robinson.[4] The transcriptions will mainly be completed by the paid research fellows and doctoral researchers who form part of the *Estoria* project team, but we are also making use of transcriptions carried out by crowdsourced volunteers. The paper will outline the specific strategies in use by the project with a hope to making crowdsourcing successful in this and later, similar projects.

2 Crowdsourcing: definition and history

Although the term is of the twenty-first century, a portmanteau of "crowd" and "outsourcing" widely attributed to Jeff Howe (2006), crowdsourcing is not a modern phenomenon. An often-cited example from history is the first Longitude Prize, a sum of £20,000—around £2 million in today's money—awarded in 1765 by the British Government to John Harrison for the invention of his marine chronometer. This was an instrument that could accurately measure a ship's longitude and therefore help to ensure a safe crossing over long distances, replacing the imprecise and sometimes fatal techniques such as astronomy, or the use of pendulum clocks which were subjected to inaccuracy caused by the motion of the ship, or changes in humidity (Betts 2006). Finding a solution to the longitude problem was essentially crowdsourced.

In his 2006 article "The Rise of Crowdsourcing", Howe describes how "technological advances are […] breaking down the cost barriers that once separated amateurs from professionals", with examples such as the rise of photo-sharing websites like iStockphoto, where amateur photographers using professional-grade equipment can sell their images for a fraction of the price of those taken by professionals. High-quality images are now widely available online, and the rules of supply and demand have forced down prices. In order to remain afloat, professionals have had to change their

---

1. Where "we" or "our" are used, they refer to the *Estoria de Espanna* Digital project team, the members of which are listed here: http://estoria.bham.ac.uk/blog/?page_id=133.

2. More information about the project can be found at http://estoria.bham.ac.uk/blog/.

3. Descriptions of the known extant manuscripts of the *Estoria de Espanna* can be found at "Manuscript Description" http://estoria.bham.ac.uk/blog/?page_id=172.

4. The reader is directed to the Textual Communities homepage http://www.textualcommunities.usask.ca/web/textual-community/home.

business models to ensure they are providing a product that amateur enthusiasts cannot. Images that *can* be provided by amateurs largely *are* provided by amateurs, meaning that the task of providing the market with basic images is now often done through crowdsourcing (Howe 2006).

In the same article, Howe goes on to explain that the labour provided by what he terms "hobbyists, part-timers and dabblers" is sometimes, but not always free, and even when it comes at a cost, is still much cheaper than paying employees using a traditional business model. He makes a link between outsourcing and crowdsourcing, but stresses the difference between the two: the former involves relocating contracting in order to make financial savings, particularly regarding labour costs, but remaining within the professional labour market, whilst the latter is similar in its outcome of financial savings through cutting labour costs, but involves tapping into the skills, expertise and enthusiasm of amateurs (Howe 2006).

2.1 *Types of crowdsourcing*

Howe (2008) has also outlined four main types of crowdsourcing: raising money through crowd-funding; harnessing the knowledge of the crowd through crowd-wisdom; users voting and therefore ranking content through crowd-voting; and cases where content is generated by the crowd through crowd-creation.

Crowd-created tasks can be divided into two subsections: microtasking and macrotasking. Microtasking involves tasks that are large in volume but can be easily separated into smaller tasks to be completed using human labour, and often without a high level of skill. Typically these are tasks which require human input, and for which a computer-based solution is prohibitively difficult or expensive. An example of microtasking would be the images uploaded by amateurs to iStockphoto, described above. Macrotasking, on the other hand, still involves discrete tasks that can be done independently and later reassembled to form part of a larger task, but the distinction from microtasking is that macrotasking requires specialised skills on the part of the worker (Walsh *et al.* 2014). Crowdsourced volunteers carrying out transcriptions of handwritten manuscripts is an example of the latter, and can be seen in the *Estoria de Espanna* Digital project, as well as the many other digital humanities crowdsourced projects such as University College London's *Transcribe Bentham*[5] and the *Revealing Cooperation and Conflict Project*[6] being run by a team of eight universities including Colorado, Wyoming and Zurich.

3 Crowdsourcing and the *Estoria de Espanna* Digital project

The *Estoria de Espanna* Digital project is using crowdsourced macrotasking in that a team of volunteers are helping to transcribe some of the folios of the *Estoria* manuscripts that we are using in the current phase of the project. The task required by all transcribers, both paid and volunteers, is to read the text in the folio image and edit the bare base text in the text box below to match that of the image. The completed transcription for each folio will include XML tags for aspects of page and text layout as well as abbreviations and scribal emendations, although not all tags have to be inserted by the same transcriber. The image below shows side 16or of the manuscript Q part-way through the transcription process using the Textual Communities transcription platform:

---

5. The reader is directed to the *Transcribe Bentham* website http://blogs.ucl.ac.uk/transcribe-bentham/.
6. The reader is directed to the *Revealing Cooperation and Conflict Project* Citizen Scholar Portal http://revealingcooperationandconflict.com/citizen-scholar-portal/.

1. Q f.160r part-way through transcription[7]

As mentioned above, the primary aim of the *Estoria* project is to transcribe and collate some of the principal versions of the *Estoria de Espanna.* One of the further aims of the project is to build a research environment to allow scholars and the general public to engage with the text and context of the *Estoria de Espanna* manuscripts. It was not the original intention of the *Estoria* project to make use of crowdsourcing, but during the first year of funding we decided to begin to use crowdsourcing on a small scale to assess the feasibility of using crowdsourcing to help us achieve our project aims. This also formed part of our approach to public engagement and impact, and has allowed us to test out the use of crowdsourcing as a potential future avenue for the continuation of the project. Since then, our use of crowdsourcing has grown through the use of specific strategies which will be explained below.

It is not our intention at this stage that crowdsourced transcriptions will replace the necessity for those carried out by paid members of the project team. The complexity of the hands and of the XML tagging requirements involved mean that this is highly unlikely within the format of the project during the current funding period (this is explained more fully below). Rather, we hope that crowdsourcers will be able to speed up the transcription process, largely by taking on some of the simpler and more laborious tasks such as line-breaking, to enable us to produce more high-quality transcriptions in the lifetime of this project. This will lead to our producing a fuller and more coherent collation than would be possible without the aid of crowdsourced volunteers. It is also a serendipitous benefit that involving members of the public in this way allows us to consciously focus on and raise our level of academic impact.

To ensure our attempt to recruit and retain crowdsourcers is successful, I will examine and review some of the strategies used in similar projects, with a view to applying them to the *Estoria* project. In this paper I will take *Transcribe Bentham* as a particular case study, given its status as the first major (Moyle *et al.* 2006, 348), and now an award-winning, initiative to involve the public in manuscript transcription through the use of crowdsourcing (Causer-Terras 2014a).

---

7. The image of manuscript Q used here can be freely viewed online on the Biblioteca Digital Hispánica section of the Biblioteca Nacional de España website http://bdh-rd.bne.es/viewer.vm?id=0000134882&page=1 (MSS/5795).

3.1 *Case study: The strategies and experiences of Transcribe Bentham and their implications for the Estoria de Espanna Digital project*

3.1.1 Background to *Transcribe Bentham*

Transcribe Bentham is a digital project based at University College, London (UCL). It forms part of the wider *Bentham Project*, which was founded in 1958, with the aim of producing the new edition of the *Collected Works of Jeremy Bentham*. A major component of this work is cataloguing and transcribing the 60,000 handwritten folios of Jeremy Bentham, a philosopher and jurist (Moyle *et al.* 2011, 347-356). *Transcribe Bentham* was launched in the autumn of 2010 when some 40,000 folios had yet to be transcribed, and sought to recruit crowdsourced volunteers to work towards transcribing some of the remaining folios (Causer *et al.* 2012, 119-137). Further intended outcomes of the project involved engaging the public in academic research, with a view to opening up the work of Jeremy Bentham to new audiences through the creation of a corpus of digitized manuscripts and transcriptions, as well as providing academia with data on public participation through crowdsourcing and some of the specific technology they developed and used (Moyle *et al.* 2011, 355). The project had an initial funding period of one year, between 2010 and 2011, the first six months of which were closely examined by members of the *Transcribe Bentham* team to analyse the effectiveness of the strategies they employed to enable crowdsourcing on the intended outcomes of the project. The project's initial funding came from an AHRC scheme entitled "Digital Equipment and Database Enhancement for Impact" (DEDEFI) (Causer-Terras 2014b, 61). Smaller-scale follow-on funding was provided for the continuation of the transcription arm of the project first by UCL, and more recently the project was awarded a substantial grant from the Andrew Mellon Foundation's Scholarly Communications programme (*ibid.*, 65-67), which expired at the end of September 2014.

    Transcribe Bentham differs from the *Estoria* project in that the former makes much heavier use of crowdsourcing as a means of transcribing folios, as *Transcribe Bentham* is a crowdsourcing-based project. The *Estoria* project, on the other hand, has as its main source of transcriptions a paid team of in-house transcribers comprising graduate students and research fellows, and crowdsourcing forms a much smaller part of the project. It is, nevertheless, useful to examine the strategies employed by *Transcribe Bentham* to analyse their effectiveness, and to consider if and how such strategies could be adapted for our own needs.

    The *Transcribe Bentham* project was careful to consider several issues that would affect the effectiveness of crowdsourcing. These were:

1) The legibility of Bentham's hand, including the orthography and scribal changes he made to his work;
2) The complexity for volunteers, many of whom would be beginners in transcription, with the requirements of TEI markup;
3) The need for careful balance between human intervention for quality control purposes, which is by its nature expensive, and a wariness of deterring participation by volunteers by unrealistically high quality standards;
4) The requirement to provide a "welcoming, rewarding and addictive experience" to crowdsourced volunteers with a range of ages, backgrounds and relevant prior experiences (Moyle *et al.* 2011).

    The same issues are also ones with which the *Estoria* project team has come into contact, and in this paper I will outline the specific strategies we have taken to deal with each one.

3.1.2 The participation of "super volunteers"

During the six-month testing period of *Transcribe Bentham*, which ran from September 2010 to March 2011, around 1009 folios were transcribed by crowdsourcers. Of these, 56% were deemed to be complete, when moderated by paid project staff. Only 259 people, or 21% of those who had signed up for accounts with *Transcribe Bentham* actually did any transcriptions, of which a small minority (0.6% of those signed up, and 2.7% of all the active transcribers) did the vast majority of the work, completing 70% of the 1009 transcribed folios (Causer *et al.* 2012, 126). More recent statistics show that by July 2013 around 13% of volunteers had been active transcribers working on at least one manuscript, and only around 5% had worked on more than one manuscript (Causer-Terras 2014b, 72-73). Such findings are, on the whole, consistent with those of Rose Holley, who describes herself on her blog as a "digital library specialist" and "pioneer of crowdsourcing" (2015). For three years, Holley managed the *Australian Newspapers Digitisation Program* at the National Library of Australia, an extremely successful project on which 40,000 crowdsourced volunteers have worked (Holley 2015). Having studied closely the use and effectiveness of crowdsourcing, she writes of a small percentage of "super volunteers" in crowdsourced projects who "consistently achieve significantly larger amounts of work than everyone else" (2010). In her research, which is based on four studies of crowdsourced projects, she found this percentage of super volunteers was typically around 10%. *Transcribe Bentham*'s statistics for 8 September 2010 to 19 July 2013 show that a smaller percentage of volunteers can be considered super volunteers, as the majority of transcription work has been carried out by seventeen individuals, which constitutes around 0.5% of the 2934 transcription account holders, or 4.5% of the 382 volunteers who had worked on at least one manuscript (Causer-Terras 2014b, 72-73). The *Estoria* project has far fewer crowdsourcers. At the time of writing (February 2015) we have around fifty volunteer transcribers signed up, of whom seven have actively been involved in transcribing folios at various levels. Of these seven, four do the vast majority of the transcriptions. Some volunteers add in line-break tags only, others also edit the bare text in the transcription to match the folio image, whilst others use full XML markup to transcribe. At the present time it can be seen that even though the *Estoria* project has far fewer volunteers than some other collaborative transcription projects, we see a percentage of super volunteers that is largely consistent with Holley's conclusions.

One strategy to raise motivation amongst volunteers in crowdsourcing projects is to use competition. The *OldWeather* project allows transcribers of ships' logs to earn points and be promoted through the ranks from cadet to captain,[8] and top contributors in the British Library's *Georeferencer* project are ranked in a leader board according to the points they have earned.[9] *Transcribe Bentham* has a points system where completing different activities or tasks for the project enable users to earn points and progress from being a "probationer" (0 points) to "prodigy" (75,000 points). There is also a leader board of those with the most points. There is no doubt that the use of competition is an effective motivator for some volunteers and would encourage them to do more transcriptions. They also have the "Benthamometer" where users can view the progress of the project to date, reinforcing the idea that all volunteers are working together towards a common goal, which motivates those who are more community-minded.[10] At the current time, the *Estoria* project does not make use of such strategies. The reason for the lack of competition between users

---

8. Readers are encouraged to visit the *OldWeather* project website at http://www.oldweather.org/.

9. Various charts and tables relating to the points of top contributors in the British Library's *Georeferencing* project can be seen at http://www.bl.uk/maps/georeferencingdata.html.

10. Readers are encouraged to visit the *Transcribe Bentham* transcription desk showing the leader board and Benthamometer at http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham.

is largely one of funding restrictions: we simply do not have the funding to enable us the time or the personnel to develop the tools which would allow us to award points to users automatically, and we do not have the time to spend awarding and counting up points for each transcriber manually. It is, however, something that we may consider working on in the future if funding allows, and if we have sufficient numbers of active crowdsourcers as to make it a worthwhile task. A progress chart like the Benthamometer would be simpler to put into practice as it could quite easily be manually updated periodically, and would be an effective way of motivating transcribers. We are in the process of discussing such a strategy and are working towards getting it up and running as part of the project blog.

3.1.3 Quality control and acknowledgement of volunteers' work

In order that the transcriptions would be of use within the digitized corpus, paid members of the *Transcribe Bentham* team were required to spend a great deal of time checking and providing quality control of the work carried out by volunteers. This is due to the complexity of Bentham's handwriting, his idiosyncratic use of language, and the technological expertise required to markup the text appropriately. Naturally, this time comes at a cost (Causer-Terras 2014b, 74-80). As Kent Anderson points out when writing of the *Transcribe Bentham* project, "even when the labor is free, the expenses incurred to coordinate and manage it well can be significant" (Anderson 2011). This is bitingly remarked upon in the comments left by readers of another blog on the subject of the expenses involved in crowdsourced transcription projects, where Bob Hillery notes that "there ain't no such thing as a free lunch" (Zou 2011). Such a comment, although self-consciously and knowingly flippant, refers to the essential cost of moderating submissions to ensure standards in crowdsourced transcription projects, and that such costs cannot be avoided if academic rigour, and by extension the later usefulness of such transcriptions are to be safeguarded. *Transcribe Bentham* has described transcription moderation by staff-members, although expensive and time-consuming, as "indispensible," and they state that feedback to volunteers and a level of moderation for quality control were "important not only to maintain the pace and quality of transcription, but they were also a vital part of the general user experience" (Causer *et al.* 2012, 127-130).

   Linked to the need for quality control of crowdsourced transcriptions, like Holley, those behind *Transcribe Bentham* stress the need to acknowledge and reward the work of crowdsourcers to maintain enthusiasm and avoid a sense that volunteers are being exploited (Causer *et al.* 2012, 130-131). This can be done through rewards such as certificates, or even by simply acknowledging the work of volunteers by naming them on project websites or inviting them to meet members of the project staff (Holley 2010). In the same vein, and particularly when working on macrotasks requiring special skills or a high level of concentration and therefore personal investment by volunteers, the *Transcribe Bentham* team highlight the importance of building on Holley's "online environment of camaraderie" such as in forums, by ensuring transcribers feel trusted, respected, and that they have any enquiries answered in a timely fashion (Causer *et al.* 2012, 130-131).

   The *Estoria de Espanna* Digital project also recognises the need for moderation and recognition of any transcriptions carried out by crowdsourced volunteers, and we are trialling various approaches to achieve this goal. At the moment, the small number of active crowdsourcers involved in the project means we can pair paid staff with volunteers, so the volunteer transcribes as much as they can, and their work is followed folio-by-folio by a staff-member who is then able to give personalised, formative feedback and praise to the crowdsourcer as appropriate. If and when our numbers of crowdsourcers rise, it is to be hoped that such a strategy would be untenable, as this would mean we have sufficient volunteer transcribers to ensure that one-to-one pairing is impossible. When this is the situation we will allocate one individual paid transcriber as the

moderator of all crowdsourced transcriptions, as far as this is feasible in terms of workload. This team-member will then be relieved of his or her other transcribing duties to afford them the time to moderate and feedback on crowdsourced folios in a timely and encouraging manner in order to acknowledge and reward volunteers, as well as to provide formative comments to improve their work where necessary and appropriate.

Further acknowledgement of volunteers' work has already been given by *Estoria* team-members' personally thanking crowdsourcers by email and using the bulletin board forum facility of Textual Communities, and we have recently started recognising the work of all transcribers, both volunteers and paid, through the "Transcriber of the Week" award on our blog.[11] The decision was taken not to restrict this recognition to crowdsourcers in order to strengthen the feeling that all transcribers form part of the same transcription team, and therefore reducing hierarchies between paid transcribers and volunteers. *Estoria* crowdsourcers have been invited to the project's annual colloquia, and their input was actively sought for the background to this paper in order to formally recognise and reward their importance to the project, as well as to help us to improve the way we handle crowdsourcers with the eventual aim of increased recruitment and retention. Volunteer transcribers gave feedback that working on the *Estoria* project was for some a welcome opportunity to return to academic-style work, and for others a way to increase their knowledge and understanding of palaeography and medieval texts and cultures. Others wrote that the interest for them was in the digital aspect of the project, coupled with the chance to have their work formally acknowledged in the eventual edition.

When asked about difficulties they had faced in working on the project, a common response was related to the macrotasking nature of transcribing and encoding, that is to say the time, effort and high level of attention required for such an intricate task. As the number of crowdsourcers involved in the project grows, we will consider the possibility of other forms of recognition, such as written and tangible acknowledgement of their input in the form of certificates of participation or commendation, and continuing to invite them to future colloquia. Furthermore, one of the founding principles of the Textual Communities transcribing platform is that all work carried out is acknowledged, and in practice, all transcribed folios show who worked on them (or who clicked save at least once whilst logged in). This means that the names of all crowdsourcers will eventually appear on the project website and in the list of contributors to the digital edition, unless they choose to opt out of this recognition process.

It is worth remarking here that *Transcribe Bentham* have reported finding that the necessary quality-control process of checking volunteer transcribed manuscripts is gaining in efficiency, with the average length of time spent by project staff checking a transcription being six minutes–far quicker than the same member of staff could transcribe and encode the text (Causer-Terras 2014b, 77-80).

At the *Estoria* project we have discerned a similar difference between the time taken for a folio to be transcribed by paid staff-members and for it to be transcribed by one of our four super volunteers and checked by staff. To take a specific example, a paid staff member transcribing and encoding one side of a folio of the manuscript Q–a manuscript with a particularly high number of abbreviations requiring XML tagging–takes on average 81 minutes; the same member of staff checking the transcription and encoding of one side by one individual crowdsourcer also working on Q takes closer to twenty-three minutes. (The crowdsourcer in question was a beginner in medieval palaeography around eight months ago, a complete novice in XML tagging when he joined the project six months ago, and all his XML training has come as part of the *Estoria* project online

---

11. Readers are referred to view the relevant webpage at http://estoria.bham.ac.uk/blog/?page_id=444.

course and individual feedback and monitoring by team-members.) The time required for quality control is similar for the other three super volunteers, but not for all crowdsourcers, as varying levels of experience and expertise in palaeography mean that naturally some make more mistakes than others. It has already been seen within the *Estoria* project that when crowdsourcers receive high-quality, timely feedback to help them to improve, the time invested in doing this benefits the project long-term—volunteer-transcribed and encoded folios can be checked by *Estoria* staff members in around a third of the time it takes to transcribe them. 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.

As mentioned above, the aspiration behind the use of crowdsourcing in the *Estoria* project is not to replace paid transcribers entirely, and it is highly unlikely that this would ever be possible, given the need for transcriptions to be done according to strict guidelines to ensure continuity between tags and expansions of abbreviations across all folios. Rather, the hope is that firstly crowdsourcing will allow us to inspire interested yet amateur members of the public to engage with the text and the transcription techniques to enhance our academic impact, and secondly that we can encourage scholars, including established academics and undergraduate and postgraduate students worldwide, to make use of the work we are doing on this project for training and research purposes covering palaeography, XML and digital humanities. 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.

3.1.4 Issues regarding marketing and the legibility of the script involved

The opportunity to sign up as a volunteer transcriber on both *Transcribe Bentham* and the *Estoria de Espanna* Digital project is open to the public via the Internet, and no qualifications or pre-requisites are required by either project. This is what makes crowdsourced projects special, as anyone can sign up, regardless of prior qualifications or experience, and many volunteers can and do get involved in such projects within academia as a learning experience. My own previous experience as a participant in a transcription MOOC (massive open online course) showed anecdotally, at least, that many of the most enthusiastic transcribers were those with no prior experience or relevant qualifications, and it was the democratised environment of online anonymity, apart from a short optional user profile, that encouraged the original participation of many such people, whereas they may have been discouraged from taking part in a more traditional academic context.[12] As crowdsourced volunteers in the *Estoria* project gain in confidence and experience, we have found that the level of anonymity begins to fall away, to the extent that at the time of writing, around ten months after the official launch of crowdsourcing within the project, we are now starting to get to know our four super transcribers quite well through personal communication in the form of moderation and feedback emails, as well as interactions on the bulletin board forum of Textual Communities. Of course, if numbers of crowdsourcers were to rise significantly in future then we may not be able to maintain these relationships with all volunteers, but we would hope to still be able to know personally our 5-10% of super volunteers.

---

12. This MOOC, entitled "Deciphering Secrets: Unlocking the Manuscripts of Medieval Spain", led by Dr Roger Martínez-Dávila (University of Colorado at Colorado Springs) formed part of the aforementioned Revealing Co-operation and Conflict project.

The requirement for special skills and levels of concentration in macrotasked activities such as transcribing historical documents means that those who do become volunteer transcribers tend to share certain characteristics. Clay Shirky (2010) has written extensively of the "cognitive surplus" of many volunteers on crowdsourced projects. This cognitive surplus can manifest itself as a desire to learn from projects such as *Transcribe Bentham* and the *Estoria de Espanna* Digital project, or as an opportunity to use their cognitive surplus to share skills from previous experience in order to contribute to projects they deem beneficial for the wider good.

Both *Transcribe Bentham* and the *Estoria* project have actively marketed their projects to certain groups in the search for volunteers. *Transcribe Bentham*, as a much more mature and heavily crowdsourcing-orientated project, has marketed more strongly than the *Estoria de Espanna* has to date. The former started out by marketing to schools providing A Level courses which include work by Jeremy Bentham on their syllabuses, to the academic sector, including those who teach and research palaeography and research methods within the humanities, and to amateur historians and enthusiasts of the subject matter of the project within the wider public (Moyle *et al.* 2011, 354). There was also an article in the *New York Times* about *Transcribe Bentham* during the Christmas season of their six-month testing period, which saw numbers of transcribers rise significantly (Causer *et al.* 2012, 125-6), and later in 2011 a *Sunday Times* article mentioned *Transcribe Bentham*, which had a similar effect on raising numbers of crowdsourcers (Causer-Terras 2014b, 67). The *Estoria de Espanna* Digital project, on the other hand, has not marketed to such a great extent. We have primarily marketed more directly to academics and postgraduates within the circles of medieval editing, medieval Iberian studies and to a lesser extent, more general medieval studies. We have also marketed to amateur palaeography enthusiasts on a small scale through forums related to the aforementioned MOOC on the topic of a similar project to ours.

The reasons for our much smaller marketing strategy than that of a project focusing more intently on crowdsourced transcriptions, such as *Transcribe Bentham* are two-fold. Firstly, if we were to market as heavily as *Transcribe Bentham* did, and our marketing were to be as successful as it was for them, it is unlikely, within the current time and budgetary constraints, that staff moderation could keep pace with volunteer contributions. This would both overwhelm staff-members on the project, and the delays in moderation and feedback to volunteers that would inevitably ensue would almost certainly discourage crowdsourcers from continuing to transcribe. This was the case in the *Transcribe Bentham* project during the Christmas break of 2010, when staff annual leave and bank holidays, followed by a period of time in January spent catching up on a backlog of volunteer transcriptions caused by the publication of the *New York Times* article during the holidays, meant there were delays of up to three weeks in moderating and feeding back on transcriptions. This delay led to a fall in the enthusiasm of some volunteers who signed up in the days directly following the *New York Times* article, and then to the project's missing out on recruiting some of these crowdsourcers as regular transcribers (Causer *et al.* 2012, 129-130). Secondly, the *Estoria* project has not marketed as heavily as *Transcribe Bentham* as it is recognised that we need any volunteer transcribers to have at least a basic background in medieval palaeography and in reading medieval Castilian before they are able to access the scripts required to complete even the most basic of transcribing tasks, namely inserting line break tags into the bare text transcriptions.

A benefit for the inexperienced palaeographer, however, is that the *Estoria* project does not require transcribers to transcribe from scratch, but rather to edit existing base text transcriptions to match, as closely as possible, how the text appears in the folio images. The script in *Transcribe Bentham* can also be described as a difficult one for the uninitiated to read, which is made even more difficult by the fact that Bentham's handwriting deteriorated as he aged, and there are many examples of interlineal additions, deletions, marginal notes, non-standard spellings, and

words in languages other than English (Moyle *et al.* 2011, 350). *Transcribe Bentham* was able to remove some of the complexity of the transcription task by allowing users to choose a folio that they considered manageable (*ibid.*, 350), by means of a basic 'difficulty' classification based on the date of the manuscript's composition. The *Estoria* project is not currently in a position to be able to allow volunteers to choose folios to transcribe because of the way in which our online transcription platform works, as specific transcribers have to be assigned a folio by the principal project investigator. This enables that particular transcriber to save any work they do on that folio, and is designed to avoid multiple transcribers accidentally working on the same folio and duplicating work. However, the characteristics of the *Estoria* manuscripts mean that transcribers choosing folios would be largely unnecessary, as the majority of folios within a given manuscript are very consistent in difficulty and layout, except for the occasional folio with a significant amount of marginalia or scribal changes, or damaged folios close to the start of each codex. It is entirely possible to assign these folios only to experienced transcribers or paid members of staff, removing the need for crowdsourcers to choose their own folios.

3.1.5 Issues regarding the complexity of tagging

One of the main issues that has faced the *Estoria de Espanna* Digital project, like *Transcribe Bentham*, is the need for volunteer transcribers to be familiar, if not with the full complexity of the Text Encoding Initiative's (TEI) XML markup, at least with the specific tags used by each project when transcribing folios. Tagging is a vital part of the transcription process of both projects, if folios are to be of use for their intended purpose once transcribed. However, the necessity to include XML tags in transcriptions, although for the most part not actually that complex in the context of the transcriptions required given that the majority of tags needed are taken from a relatively small pool of tags, appears daunting to the unitiated and could be off-putting to potential transcribers. The *Transcribe Bentham* project has dealt with this issue by the use of a transcription toolbar. This is a series of buttons available to transcribers which, when clicked, automatically input the relevant TEI tag into the transcription (Moyle *et al.* 2011, p352-3). Examples include line-break, ampersand and marginal note. Such a tool removes the necessity for volunteers to learn to tag from scratch using XML, and therefore lowers the complexity of the transcription process with a view to creating a seemingly more manageable task, and encouraging more participation by a wider range of volunteers.

The *Estoria* project has taken a different approach to dealing with the complexity of tagging. Some of the tags required are very complicated, even for confident transcribers, such as the tags for scribal emendations, but many tags are relatively simple and are repeated multiple times in almost every folio. An example of the latter would be our expansion of "q-macron", for the word "q*ue*" when the letters "ue" are replaced by the abbreviating macron:


2. Q f.126r.

q<am>¯</am><ex>ue</ex>

Here, the abbreviation mark ("am") is expanded ("ex") to the letters "ue". A list of common tags required by the folios of the *Estoria de Espanna*, including examples from the manuscripts,

is freely available for transcribers, both crowdsourced and paid, in the form of a wiki page entitled "Transcription Guidelines" within Textual Communities.[13] Crowdsourced transcribers have made, and continue to make use of this when transcribing. The vast majority of tags required for transcribing on this project are included in this wiki page. However, the difficulty with beginners working primarily from a pre-prepared list of tags in the particular case of the *Estoria de Espanna* manuscripts is that the sometimes-inconsistent abbreviations within and between manuscripts require transcribers to edit the XML tags to fit the specific abbreviation they require. For example, the majority of abbreviations of the word "grand" in the manuscript E2 appear as "gn̄d" and are tagged as follows:

<choice><abbr>gn<am>¯</am>d</abbr><expan>g<ex>ra</ex>nd</expan></choice>

This relatively simple abbreviation mark requires quite a complex tag. The first part of the tag (<abbr>, that is to say 'abbreviation') shows that the letters g n d appear, and that the n has an abbreviation mark above it, which we consider to be a macron. The second part of the tag (<expan> – 'expansion') shows that the macron represents the abbreviated letters "ra", and shows where the missing letters should be expanded. The expanded form will then display in the eventual electronic edition as "*gra*nd", with the "ra" italicised to show they have been expanded and do not appear in the original manuscript image. The above tag appears in the Transcription Guidelines as the appropriate tag for use with "grand", when it appears like this.

There are several examples, however, of where "grand" is written differently, particularly in other manuscripts, but even within E2—often in the folios of E2 inserted in the fourteenth century to fill in gaps. If the abbreviation appears differently, the above tag would be incorrect. For example, the abbreviation shown here, taken from Q folio 47v, shows the letters g r a t with an abbreviating macron above the letter a.[14]
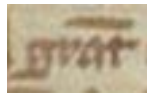
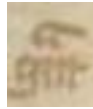3. Q f.47v #1

This word should be tagged as follows:

gra<am>¯</am><ex>n</ex>t

to show that the abbreviation mark (<am>) represents the letter n, and that in the expanded form (<ex>) this should appear between the a and the t, as "gra*n*t". This also shows an orthographic variation in the word, from "gra*n*d̲" to "gra*n*t̲".

In the same folio of Q we see the word abbreviated in the following way:

---

4. Q f.47v #2

Here we have the letters g n t with a superscript a above the n. The tag would therefore be: <choice><abbr>gn<am>ᵃ</am>t</abbr><expan>g<ex>ra</ex>nt</expan></choice> and the word would appear in the edited transcription as "*gra*nt".

In both of these cases, the tag included in the wiki for "grand" would clearly not represent the word as it appears in the manuscript images, yet it is unfeasible to include a list of every single tag that may be required by transcribers in order that they could identify and then copy and paste the one they need into their transcription in one document. At any rate, the compilation of such a list would necessitate the reading of every single abbreviation in every single folio, which is clearly impossible within time and budgetary constraints, simply to include these in transcription training tools to allow transcribers to copy and paste every tag. Because of the need to edit tags in this way, a transcription tool where users can automatically input tags would not be possible for the *Estoria de Espanna* project for anything but the simplest tags, and therefore the level of complexity of tagging cannot be relieved in this way. Similarly, if transcribers are to be able to accurately tag abbreviations and expansions, they must fully understand the formation of tags in order to be able to edit the examples included in the wiki.

4 Crowdsourcing and the *Estoria de Espanna* Digital project: XML as a recruitment barrier and the need to address this

The final section of this paper will deal with the issue of the comprehension of XML as a recruitment barrier necessitating specific training tools for potential crowdsourcers to ensure the success of crowdsourcing in the *Estoria de Espanna* Digital project. The understanding of basic XML was recognised as a stumbling block in the recruitment of crowdsourced volunteers to the project in the early days after the launch of crowdsourcing in April 2014, as many of the experienced palaeographers within medieval Hispanic studies who we approached as potential crowdsourcers appeared somewhat reticent to continue transcribing or to publicise the possibility of becoming crowdsourcers to their graduate students, potentially because of a lack of knowledge of XML.

As a strategy towards overcoming the barrier of an understanding of basic XML tagging, in order to encourage experienced and intermediate palaeographers to sign up as crowdsourced transcribers, two of the graduate students on the *Estoria de Espanna* Digital project team have developed an online course.[15] The course uses Canvas, the online open-source learning management system used by the University of Birmingham, and has been designed in such a way that all features are fully available for use by non-members of the university, without the need for a Birmingham academic email account. The course is available in both English and Spanish, and has been planned to enable the user to move through the material at his or her own pace, starting with

---

15. The graduate students who developed the course are Christian Kusi-Obodum and myself (both University of Birmingham). The course was translated into Spanish by Dr Enrique Jerez Cabrero (University of Birmingham) and Alicia Montero Málaga (Universidad Autónoma de Madrid). The online course can be viewed here: https://canvas.bham.ac.uk/courses/6673.

the most basic of tagging elements and moving to complex tagging issues. A new crowdsourcer can start at the most basic level of choosing which of three line break tags is the most appropriate, which can all be copied and pasted without the need for any editing of tags. If the crowdsourcer chooses to do so, they can end or pause their learning at this stage and request to be assigned folios to which to add line breaks. Alternatively, the volunteer transcriber can work their way through each course module, which increase in complexity, until they have covered all types of tags that they may need to utilise when participating in the *Estoria* project. Crowdsourcers can return to any part of the course at will to brush up on the particular tagging issue required by their folio. For example, if a transcriber meets a scribal change for the first time they can return to the scribal changes module of the course to refresh their memory. There is also a frequently asked questions section of the course, and users are encouraged to ask any questions not answered, or which they feel have been left unclear, on the bulletin board forum of Textual Communities. Although crowdsourcing at the *Estoria* project is still in its relative infancy, we have already seen evidence of volunteers using the course through their understanding of tagging in their transcriptions, and also by their asking relevant and thoughtful questions on the forum. As Holley (2010) advocates, we are working hard to ensure that the bulletin board is an environment to foster a sense of camaraderie amongst volunteers and staff alike, and to ensure that crowdsourcers feel valued members of the team.

The course is structured into the following modules:

1) Introduction to crowdsourcing

2) Configuring your machine and working with XML

3) Putting line-breaks into a folio

4) Text structure (column boundaries, anonymous blocks and divisions)

5) Proofing the base text (making the transcription text match the text in the image, before XML encoding takes place)

6) Basic tagging

7) "Choice" tags

8) Marginalia and scribal changes

9) Tagging split rubrics

10) Final stages of a transcribed folio (parsing)

11) Frequently asked questions

The modules included in the course mirror the information included in the Transcription Guidelines wiki, with the main difference being that the course has actively been designed with accessibility of material in mind for those who are new to transcribing, and/or to the use of XML. The course is not meant to replace the need for the Transcription Guidelines, but rather to complement it, and to fulfil the training needs of an inexperienced transcriber in order to encourage their confidence and enthusiasm in the project. The course does not aim to include all of the information in the wiki, although it does introduce all of the main issues involved in transcribing for the *Estoria* project, and on more than one occasion course-users are directed to consult the wiki page for supplementary information if required. Experienced palaeographers and users of XML who sign up as volunteer transcribers have the option to bypass the course entirely, and to transcribe using just the information contained in the Transcription Guidelines, or they can cherry-pick the

sections of the course they feel they need. These are wholly appropriate possibilities within the differentiated needs of the range of crowdsourcers who may choose to join the project as volunteer transcribers.

In line with the well-known and used three-part pedagogic structure, each module starts with a clear outline of what will be studied and the learning outcomes students should achieve therein. Following the pedagogical concept of "chunking", the content of the module is then divided into manageable sections, with any subject-specific technical language explained as plainly as possible and, where appropriate, images from the manuscript are included as examples. The learning outcomes for each module are then checked in a plenary section, to enable trainee transcribers to self-assess their learning for that section and judge whether to move onto the next section of the course or to repeat the module. Sometimes these plenary sections are very simple and merely recap what the transcriber should know having completed the module, and on other occasions the plenaries are more elaborate and include the use of quizzes. Also included in the course is a short screencast video of a member of the *Estoria* team transcribing using the feature just taught in the course, while narrating and explaining her tagging choices for the benefit of the learner.

Despite it only being four months since the online course was launched in November 2014, we have already seen evidence of its success. When monitoring crowdsourcers' transcriptions and providing feedback, if a volunteer is consistently making the same mistake, 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. If further clarification is needed, the staff-member can, of course, provide specific and tailored feedback or explanations. Being able to direct volunteers to the course allows us to cut down the time, and therefore expense, involved in providing crowdsourcers with mentoring of transcribing at the more basic levels. A further benefit has been seen in that due to the "chunked" nature of the content in the course, new volunteers are less likely to feel overwhelmed at the prospect of starting to tag transcriptions using XML for the first time. One particular crowdsourcer joined the project as a line-breaker, which is the most basic of levels for volunteer transcribers and, as a complete novice to XML, was at first reluctant to attempt any more complex tagging. Sticking to line-breaking is a perfectly valid choice for a new crowdsourcer in the *Estoria* project. To line-break successfully he only had to complete modules one to three of the online course. After several folios of line-breaking successfully, the member of staff providing quality control and one-to-one feedback of his work suggested he try the next levels of crowdsourcing, which is proofing the base text and including some simple abbreviation expansion tags, and directed him to modules four to six of the course. The crowdsourcer found he no longer felt daunted at the prospect of tagging using XML and decided to do all remaining modules. This, coupled with ongoing one-to-one mentoring and support has meant that this transcriber is now one of the most prolific and accurate of our four super volunteers.

5 Conclusions and aspirations for crowdsourcing at the *Estoria de Espanna* Digital project

In conclusion, examining the strategies aiming for the success of crowdsourcing in the *Transcribe Bentham* project, and reviewing their adaptation and application within the *Estoria de Espanna* Digital project reveals that the theory behind recruiting and retaining volunteer transcribers is simple. It is necessary, however, for projects to ensure they adapt this theory to fit their own requirements and capabilities, within the confines of their resources, if they wish for similar strategies to be successful within their own context. To summarise the findings of this paper, in order to recruit crowdsourcers, projects must specifically target their marketing and bear in mind

the restrictions of the time and expense available for the necessary moderation and quality control of volunteer-produced transcriptions. They must also provide individual feedback to volunteers, within the bounds of the particular project, as marketing too widely or at inopportune moments can be detrimental to the retention of newly-recruited crowdsourcers, since delays in moderation or feedback can cause volunteers to lose interest. Whilst collaborative transcription projects seeking to make use of crowdsourcing as a method of producing transcriptions must accept that doing so will necessitate a certain level of time-input and expense in the form of quality control and feedback, there are ways in which projects can mitigate against such demands and streamline the process, such as the use of online tutorials. Furthermore, as individual transcribers become more experienced the requirement for mentoring or lengthy one-to-one explanatory feedback does reduce, and the quality of their transcriptions does rise, meaning that the process becomes more efficient as volunteers become more confident. Projects should also be aware that to retain volunteers they should be ready to continue to challenge them, in the light of Shirky's (2010) research into the cognitive surplus of those who take part in more demanding crowdsourced activities of this type: *Transcribe Bentham*'s approach to this is to allow crowdsourcers to choose their own folios to transcribe, so those wanting more challenge can choose more difficult folios, whilst the *Estoria* project online training system allows volunteers to start working at a basic level and move to more challenging tagging as they gain confidence, which continues to feed their cognitive surplus. Rewards and acknowledgement of volunteers' work is also fundamental to retaining volunteers, as Holley (2010) shows, and there are various methods of showing crowdsourcers that their efforts are valued, such as feedback, awards and certificates. A further important point of Holley's (*ibid.*) that those behind collaborative transcription projects such as *Transcribe Bentham* and the *Estoria* project must be aware of and accept is that not all transcription account-holders will be active transcribers, and only around 10% (as an average maximum) of active transcribers will produce the vast majority of transcriptions—that is to say to be super volunteers.

As mentioned above, the *Estoria de Espanna* Digital project is still in the early days of using crowdsourcing to produce transcriptions, and we are not a crowdsourcing-based project as *Transcribe Bentham* is. We do, however, hope to continue using crowdsourcing as a method of both speeding up the transcription process—we can see evidence that this is starting to be the case—and therefore allowing us to produce a fuller and more coherent eventual collation. Furthermore, using crowdsourcing is a useful way of enabling us to continue to improve our level of academic impact by harnessing the enthusiasm of the "hobbyists, part-timers and dabblers", to use Howe's aforementioned phrase (2006).

Within the current funding phase it is not the intention of the *Estoria* project to replace paid transcribers with volunteers, but rather to enhance the way in which we produce transcriptions, whether this is by moderating some folios more quickly than we could transcribe them ourselves, or by having volunteers lessen the load of transcribing by line-breaking and proofing the folio prior to the more complex transcription tasks such as encoding, which can be carried out by paid team-members. It is possible that future sources of funding may allow us to concentrate more directly on crowdsourcing, giving it a larger emphasis within the project, and we are currently looking into possible avenues for this. Routes we may take include developing a MOOC, which would allow us to target more potential crowdsourcers, and employing a member of staff who would have at least a proportion of their time specifically dedicated to issues of crowdsourcing. This would allow us to market more widely, but doing so may also have practical implications as the level of one-to-one feedback we currently give our volunteers may be unfeasible if numbers of crowdsourcers were to rise significantly, particularly if the MOOC were to be popular. We are also looking into the potential usefulness of a WYSIWYG (what you see is what you get) transcription toolbar along the

lines of that of *Transcribe Bentham*, which would input some of the most common XML tags at the touch of a button, although, as explained earlier, this would require us to research its implications, bearing in mind the inconsistent nature of some of the abbreviations included in the manuscripts we are transcribing. The fact remains that crowdsourcing for collaborative transcription projects such as *Transcribe Bentham* and the *Estoria* project is an upcoming and exciting area of methodology and research within digital humanities, which offers many possibilities for future study.

6 **Acknowledgements**

7 Works Cited

Anderson, Kent, 2011, "Even crowdsourcing can get too expensive", *The Scholarly Kitchen*, 14-03-2011, http://ow.ly/JFXzj (checked 26-09-2014).

Betts, Jonathan, 2006, *John Harrison (1693–1776) and Lt. Cdr Rupert T. Gould R.N. (1890–1948)*, digitally self-published. http://www.rmg.co.uk/sites/default/files/media/pdf//Gould-Harrison-longitude-JBetts.pdf (checked 26-02-2015).

Causer, Tim; Terras, Melissa, 2014a "Crowdsourcing Bentham: beyond the traditional boundaries of academic history", *International Journal of Humanities and Arts Computing* 8 (1), 46-64. http://dx.doi.org/10.3366/ijhac.2014.0119

Causer, Tim; Terras, Melissa, 2014b, "'Many hands make light work. Many hands together make merry work': *Transcribe Bentham* and crowdsourcing manuscript collections" in *Crowdsourcing our cultural heritage*, Mia Ridge ed., Farnham, Ashgate, 57-88. http://dx.doi.org/10.1093/llc/fqs004

Causer, Tim; Tonra, Justin; Wallace, Valerie, 2012, "Transcription maximized; expense minimized? Crowdsourcing and editing *The collected works of Jeremy Bentham*", *Literary and Linguistic Computing* 27:2, 119-137. http://dx.doi.org/10.1093/llc/fqs004

Zou, Jie Jenny, 2011, "Civil war project shows pros and cons of crowdsourcing", *Wired Campus*, 14-06-2011 (updated 21-06-2011), http://ow.ly/JFXX5 (checked 26-09-2014).

Holley, Rose, 2015, "Blogger profile", *Rose Holley's Blog*, http://ow.ly/OgG0k (checked 13-06-2015).

Holley, Rose, 2010, "Crowdsourcing: how and why should libraries do it?", *D-Lib Magazine* 16 (3/4). http://dx.doi.org/10.1045/march2010-holley

Howe, Jeff, 2008, *Crowdsourcing: Why the power of the crowd is driving the future of business*, New York, Crown Publishing Group.

Howe, Jeff, 2006, "The rise of crowdsourcing", *Wired* 14-06-2006, http://ow.ly/JFY52 (checked 26-09-2014).

Moyle, Martin; Tonra, Justin; Wallace, Valerie, 2011, "Manuscript transcription by crowdsourcing: *Transcribe Bentham*", *Liber Quarterly* 20 (3/4), 347-356. http://dx.doi.org/10.1045/march2010-holley

Shirky, Clay, 2010, *Cognitive surplus: creativity and generosity in a connected age*, London, Penguin.

Walsh, Brandon; Maiers, Claire; Nally, Gwen; Boggs, Jeremy; Praxis Program Team, 2014, "Crowdsourcing individual interpretations: Between microtasking and macrotasking", *Literary and Linguistic Computing*, 379-386. http://dx.doi.org/ 10.1093/llc/fqu030

Websites referred to in the text

Estoria de Espanna Digital project blog
http://estoria.bham.ac.uk/blog/

Estoria de Espanna Digital project transcription guidelines
http://www.textualcommunities.usask.ca/web/estoria-de-espanna/wiki/-/wiki/Main/Transcription+Guidelines

Estoria de Espanna Digital project online transcription training course
https://canvas.bham.ac.uk/courses/6673

Textual Communities online transcription platform
http://textualcommunities.usask.ca/api/client/index.html?community=1

Transcribe Bentham project blog
http://blogs.ucl.ac.uk/transcribe-bentham/

Transcribe Bentham transcription desk
http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham

Revealing Cooperation and Conflict Citizen scholar portal
http://revealingcooperationandconflict.com/citizen-scholar-portal/

Biblioteca Nacional de España website – images of Manuscript Q (MSS/5795)
http://bdh-rd.bne.es/viewer.vm?id=0000134882&page=1

OldWeather project
http://www.oldweather.org/

British Library's *Georeferencing* project contributor data
http://www.bl.uk/maps/georeferencingdata.html