tranScriptorium

D7.2: Second Report on Dissemination: Report on Task 7.1 (Year 2)

UPVLC, UIBK, NCSR, UCL, INL and ULCC

Distribution: Public

tranScriptorium
ICT Project 600707 Deliverable D7.2

December 31, 2014

Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development.

Project ref no.	ICT-600707
Project acronym	**tranScriptorium**
Project full title	**tranScriptorium**
Instrument	STREP
Thematic Priority	ICT-2011.8.2 ICT for access to cultural resources
Start date / duration	01 January 2013 / 36 Months

Distribution	Public
Contractual date of delivery	December 31, 2014
Actual date of delivery	December 31, 2014
Date of last update	January 17, 2015
Deliverable number	D7.2
Deliverable title	Second Report on Dissemination: Report on Task 7.1 (Year 2)
Type	Report
Status & version	Final
Number of pages	14
Contributing WP(s)	7
WP / Task responsible	UPVLC
Other contributors	UIBK, NCSR, UCL, INL and ULCC
Internal reviewer	Joan Andreu Sánchez and Tim Causer
Author(s)	UPVLC, UIBK, NCSR, UCL, INL and ULCC
EC project officer	Jose María del Águila
Keywords	Dissemination

The partners in **tranScriptorium** are:

Universitat Politècnica de València - UPVLC (Spain)
University of Innsbruck - UIBK (Austria)
National Center for Scientific Research "Demokritos" - NCSR (Greece)
University College London - UCL (UK)
Institute for Dutch Lexicology - INL (Netherlands)
University London Computer Centre - ULCC (UK)

For copies of reports, updates on project activities and other **tranScriptorium** related information, contact:

The **tranScriptorium** Project Co-ordinator
Joan Andreu Sánchez, Universitat Politècnica de València
Camí de Vera s/n. 46022 València, Spain
jandreu@prhlt.upv.es
Phone (34) 96 387 7358 - (34) 699 348 523

Copies of reports and other material can also be accessed via the project's homepage: http://transcriptorium.eu/

© 2014, The Individual Authors
No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Executive Summary

This document reports the dissemination activities for the tranScriptorium project during the second reporting period of the project. The document also proposes a dissemination plan for the forthcoming period and a preliminarly proposal for the exploitation plans for the third year.

Contents

1 Introduction **4**

2 Dissemination strategy **4**
2.1 Main target audiences 4
2.1.1 Scientific community 4
2.1.2 Content providers community 5
2.1.3 Society 6
2.2 Internal dissemination 6

3 Dissemination Activities **6**
3.1 Dissemination material 6
3.1.1 Project documentation 6
3.1.2 Project website and other online material 7
3.1.3 Prototypes and databases 7
3.2 Publications and other research dissemination activities 8
3.2.1 Publication of research results 8
3.2.2 Conferences and Contests organised by tranScriptorium members 10
3.2.3 Seminars organised by tranScriptorium members 10
3.3 Dissemination for content providers 11
3.4 Contacts and dissemination at industrial companies 12
3.5 In the media 12

4 Dissemination plan for the forthcoming period **12**

5 Exploitation plan for the forthcoming period **13**

6 Conclusion **14**

1 Introduction

The main goal of WP7 is to bring the scientific and technological results of the tranScriptorium project to the attention of our stakeholder communities in many different sciences and research fields. In this deliverable we provide an overview of all dissemination activities carried out to this effect during the second year of the project.

In deliverable D7.1 we identified the project's stakeholders in more detail. During this year we addressed the dissemination activities to these identified audiences and we have sought to identify additional stakeholders. A summary of these stakeholders is included in Section 2. Then follows a record of our specific dissemination actions carried out during the M13-M24 period. Finally we propose a dissemination plan for the forthcoming period, and an exploitation plan to be developed during year three.

2 Dissemination strategy

Dissemination activities were defined in deliverable D7.1 and they addressed three main target audiences: i) the scientific community, ii) the content provider community; iii) a third group, the general public, was identified during the course of the project. In this section we review the stakeholders within these categories that were defined in deliverable D7.1.

2.1 Main target audiences

2.1.1 Scientific community

Within the scientific community, results from the tranScriptorium project were defined as being of interest to other research groups in the fields of Crowdsourcing, Digital Humanities, Computational Linguistics, Handwritten Text Recognition, Image Processing, Document Image Analysis and Pattern Recognition.

A list of events in these areas which we considered it important to attend was enumerated in the deliverable D7.1. For this second year a list of key events was specified and a concrete programme decided upon the partners agreeing to collaborate to enable participation in these events. From this discussion, the following events held in 2014 were concretely identified as the most relevant:

- **ICFHR:** *International Conference on Frontiers in Handwriting Recognition.* This is without any doubt one of the major conferences related to Handwritten Text Recognition. Therefore, the tranScriptorium consortium agreed upon having adequate representation in 2014 edition. ICFHR 2014 was itself co-organized by NCSR, one of the tranScriptorium partners.
- **DAS:** *IAPR International Workshop on Document Analysis Systems.* This is also a relevant conference for tranScriptorium, and the project was also represented in that conference.
- **ICPR:** *International Conference on Pattern Recognition.* This is a relevant conference for Pattern Recognition in general with large impact in the PR community. The project was also represented in that conference.
- **IbPRIA:** *Iberian Conference on Pattern Recognition and Image Analysis* is also a relevant conference for Pattern Recognition in general. tranScriptorium was also present in this event.

In addition, a series of relevant international journals were identified that use to disseminate the kind of research that is carried out in tranScriptorium:

- *IEEE Trans. on Pattern Analysis and Machine Intelligence*
- *IEEE Trans. on Audio, Speech and Language Processing*
- *Pattern Recognition*
- *Pattern Recognition Letters*
- *International Journal of Document Analysis and Recognition*
- *Computational Linguistics*
- *Computer, Speech and Language*
- *Neurocomputing*

Some articles have been submitted and accepted by these journals, while other are undergoing the review process.

2.1.2 Content providers community

In this section we considered dissemination as a method to encourage the future exploitation of tranScriptorium tools, for which we identified potential end users or applications.

The clearest application of the tranScriptorium project to the content providers community is the automated transcription of historical handwritten documents. So, the main target group for dissemination and exploitation efforts are archives, libraries and other cultural heritage institutions and associations.

One of the most efficient ways to reach this community is to apply to give talks at their national and international meetings. Some of these meetings are:

- **German Archives Day.** This event takes place every year and gathers about 1000 visitors mainly from German archives but also from abroad.
- **ARCHIVISTICA.** This is the largest fair in Europe dedicated to archival issues.

In addition, an additional event was identified in 2014 to be importance in addressing the content providers community:

- **Digital Access to Textual Cultural Heritage.** This was the first time that this event was organized.

This event was organized by the Succeed project[1] and many archives, libraries and industrial representatives were due to attend. Another way of reaching this community is to make direct contact with relevant institutions (archives, libraries, and so on), and tranScriptorium took some actions along this line as we describe in Section 3.

Another interesting community to reach are developers of research infrastructure portals (in which the HTR technology could be implemented). In this case, it is necessary to contact them directly and to introduce them the technology and the results in a face-to-face meeting. Some of these portals, which were identified and contacted by tranScriptorium along the second year, were:

[1] `http://succeed-project.eu/`

- **Europeana.** Multi-lingual online collection of millions of digitised items from European museums, libraries, archives and multi-media collections.
- **TEL.** The European Library.
- **BL Labs.** The British Library Labs.
- **EHRI.** European Holocaust Research Infrastructure.
- IMPACT Centre of Competence.

tranScriptorium has kept in contact with most of the centres named above.

2.1.3 Society

In addition to the above-mentioned dissemination activities, the consortium undertook a number of activities designed to make public the results of the project. These included:

- **Website.** A dedicated project site which serves as a central repository for all reports and documentation on tranScriptorium.
- Public dissemination through the social networks.
- Publication of news in the media.

2.2 Internal dissemination

Internal dissemination was also identified as a key factor for ensuring that all partners are kept abreast of recent developments across all seven work packages. As indicated in the Description of Work, meetings have been held twice this year for this purpose. Meanwhile, day-to-day communication is achieved primarily using online communication channels, such as email and Skype. The project website and the wiki have been also used as a platform for data and information exchange among partners.

3 Dissemination Activities

In this section we give a comprehensive overview of all dissemination activities carried out in the tranScriptorium project during the M13-M24 period.

3.1 Dissemination material

3.1.1 Project documentation

In accordance with that agreed as part of the grant agreement, project documentation was developed in collaboration with all partners. A new project leaflet[2] was prepared for disseminating the tranScriptorium tools and results. This leaflet was used to disseminate the tranScriptorium results in several events like the "International Conference on Frontiers on Handwritten Recognition 2014" and the "Proposers' day" which was held at Firenze (Italy) the 9 and 10 of October 2014. It was also disseminated through several institutions like the "Biblioteca Nacional de España".

[2] http://transcriptorium.eu/wp-content/uploads/2014/12/leafletTS.pdf

3.1.2 Project website and other online material

The project website[3] is our main public dissemination tool and is easily accessible. The website presents general project information, the consortium description, scientific publications performed during the project, news about the project and public deliverables.

It also includes several prototypes and demonstrations for the technology[4] which visitors to the website can try in order to understand the research that is being developed in tranScriptorium. In addition, some of the datasets which are being used in the project (for example, see Section 3.2.2 can be downloaded through the project web site.

Website access statistics have been monitored throughout the year. Table 1 reports the statistics for the second year of project.

Table 1: tranScriptorium website access statistics

Number of visits (total)	5,154
Number of unique visitors	3,317
Number of page views	13,252
Average time on site (sec)	157

Table 2 shows from which countries the website has been accessed. It can clearly be seen that tranScriptorium website has attracted a significant number of visits originating from countries far beyond those in which the original project consortium partners are based.

Table 2: tranScriptorium website access statistics per country

Country	Number of visits	% of visits
Spain	1,331	25,82
United Kingdom	601	11,66
United States	433	8,40
Germany	293	5,68
Brazil	268	5,20
France	223	4,33
Austria	218	4,23
Greece	175	3,40
Italy	153	2,97
Netherlands	140	2,72

tranScriptorium also has profiles on the main online networks, namely, Facebook[5] and Twitter[6].

3.1.3 Prototypes and databases

Several prototypes can be accessed through the tranScriptorium web page. These prototypes show the developments both for interactive transcription and for key word spotting.

[3] http://transcriptorium.eu
[4] http://transcriptorium.eu/demonstrations/
[5] https://www.facebook.com/tranScriptorium
[6] https://twitter.com/tranScriptorium

In addition, some of the datasets which are being used in the project can be downloaded through the project web site. These databases have been prepared to allow other researchers to reproduce the experiments which are being carried out in the project and to carry out their own further research activities. The distribution policy for each database depends on the copyright status of each collection.

3.2 Publications and other research dissemination activities

3.2.1 Publication of research results

tranScriptorium has put great deal of emphasis into disseminating its research activity through the usual channels such as international conferences and journals. This is arguably the most effective channel of dissemination.

The following lists give an overview of all papers and communications done by tranScriptorium partners in 2014 (some accepted contributions for 2015 are also included).

Conferences and Workshops

1. V. Bosch-Campos, A. H. Toselli, and E. Vidal, *Semiautomatic Text Baseline Detection in Large Historical Handwritten Documents*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 690–695.

2. V. Bosch, I. Bordes-Cabrera, P. C. Muñoz, C. Hernández-Tornero, L. A. Leiva, M. Pastor, V. Romero, A. H. Toselli, and E. Vidal, *Computer-assisted Transcription of a Historical Botanical Specimen Book: Organization and Process Overview*, in Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATeCH 2014), 2014, pp. 125–130.

3. B. Gatos, G. Louloudis, T. Causer, K. Grint, V. Romero, J.A. Sánchez, A.H. Toselli, and E. Vidal. *Ground-Truth production in the tranScriptorium project*, in 11th IAPR International Workshop on Document Analysis Systems (DAS), 2014, pages 237–241.

4. B. Gatos, G. Louloudis, and N. Stamatopoulos. *Segmentation of Historical Handwritten Documents into Text Zones and Text Lines*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 464–469.

5. B. Gatos, N. Stamatopoulos, G. Louloudis, and S. Perantonis. *H-DocPro: a Document Image Processing Platform for Historical Documents*, in Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATeCH 2014), 2014, pp. 131–136.

6. D. Martín-Albo, R. Plamondon, and E. Vidal. *Training of On-line Handwriting Text Recognizers with Synthetic Text Generated Using the Kinematic Theory of Rapid Human Movements*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 543–548.

7. K. Ntirogiannis1, B. Gatos, and I. Pratikakis. *ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 809–813.

8. A. Papandreou, B. Gatos, G. Louloudis *An adaptive zoning technique for efficient word retrieval using dynamic time warping*, in Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATECH 2014), 2014, pp. 147–152.

9. I. Pratikakis, K. Zagoris, B. Gatos, G. Louloudis, and N. Stamatopoulos. *ICFHR 2014 Competition on Handwritten KeyWord Spotting (H-KWS 2014)* , in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 814–819.

10. J. Puigcerver, A. H. Toselli, and E. Vidal. *Word-Graph-Based Handwriting Keyword Spotting of Out-Of-Vocabulary Queries*, in 22nd International Conference on Pattern Recognition (ICPR), 2014, pages 2035–2040.

11. J. Puigcerver, A. H. Toselli, and E. Vidal. *Word-Graph and Character-Lattice Combination for KWS in Handwritten Documents*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 785–790.

12. J. A. Sánchez, V. Bosch, V. Romero, K. Depuydt, and J. de Does. *Handwritten Text Recognition for Historical Documents in the tranScriptorium Project*, in Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATeCH 2014), 2014, pages 111–117.

13. J.A. Sánchez, V. Romero, A.H. Toselli, and E. Vidal. *ICFHR2014 Competition on Handwritten Text Recognition on tranScriptorium Datasets (HTRtS)* in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 181–186.

14. N. Stamatopoulos, B. Gatos, and G. Louloudis. *A Novel Transcript Mapping Technique for Handwritten Document Images*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 41–46.

15. J. Tanha, D. J. Does, and K. Depuydt. *An intelligent sample selection approach to language model adaptation for hand-written text recognition*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 349–454.

16. A.H. Toselli, and E. Vidal. *Word-Graph based Handwriting Key-word Spotting: Impact of Word-Graph Size on Performance*, in 11th IAPR International Workshop on Document Analysis Systems (DAS), 2014, pages 176–180.

17. M. Villegas and A. H. Toselli. *Bleed-through Removal by Learning a Discriminative Color Channel*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 9–14.

18. K. Zagoris, I. Patrikakis, and B. Gatos. *Segmentation-based Historical Handwritten Word Spotting using Document-Specific Local Features*, in 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2014, pp. 9–14.

Publications in journals

- F. Álvaro, F. Cruz, J. A. Sánchez, O. Ramos, and J. M. Benedí, *Structure Detection and Segmentation of Documents Using 2D Stochastic Context-Free Grammars*, Neurocomputing, 150, Part A, pp. 147–154, 2014.

- A. Papandreou, B. Gatos, S.J. Perantonis, and I. Gerardis. *Efficient skew detection of printed document images based on novel combination of enhanced profiles*, International Journal of Document Analysis and Recogntion, 17, pp. 433–454, 2014.

Book chapters

- T. Causer and M. Terras, *"Many hands make light work, many hands together make merry work": Transcribe Bentham and crowdsourcing manuscript collections*, in M. Ridge (ed.), Crowdsourcing Our Cultural Heritage (Ashgate, 2014)[7].

[7] `http://discovery.ucl.uk/1393567/`

Given talks During the second year, tranScriptorium partners were invited to give several talks in organisations external to the project:

- E. Vidal. *Indexing and Searching Handwritten Text Images*, DGA workshop, Multimedia Information Processing and Information Fusion, Télécom ParisTech, Paris, 1-2 July 2014.
- P. Schofield. *Jeremy Bentham and the Computer Age: Reflections on Crowdsourcing the Transcription of Handwritten Documents*, Invited talk at ICFHR 2014, Crete, 2 September 2014.

3.2.2 Conferences and Contests organised by tranScriptorium members

tranScriptorium members were involved in the organization of conferences and contests along 2014 very relevant related to the project research.

- **International Conference on Frontiers on Handwritten Recognition 2014**[8].
  The NCSR group co-organized this conference in 2014 that was held at Greece. tranScriptorium was promoted in this conference with the help of NCSR.
- ICFHR2014 Competition on Handwritten Text Recognition on tranScriptorium Datasets (HTRtS)[9].
  This competition was organized by the UPVLC partner. The main goal in this contest was to promote the research in HTR. The Bentham dataset was used in the contest. Seven research groups were registered in the contest.
- ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)[10].
  This competition was organized by the NCSR partner. The main goal in this competition was document image binarization. Some data from the Bentham dataset was used in the contest. Seven research groups participated in the contest.
- ICFHR 2014 Competition on Handwritten KeyWord Spotting (H-KWS 2014)[11].
  This competition was organized by the NCSR partner. The main goal in this competition was to research on KWS. Some data from the Bentham dataset was used in the contest. Five research groups were registered in the contest.

3.2.3 Seminars organised by tranScriptorium members

UPVLC members were also involved in the organization of the tutorial *Handwritten Text Recognition: Word-Graphs, Keyword Spotting and Computer Assisted Transcription*[12] at the ICFHR 2014. The topic of this tutorial was related to the research in tranScriptorium. All the material provided to the participants including datasets and tools were tested in tranScriptorium. About 20 people attended to this tutorial.

[8] http://www.icfhr2014.org/
[9] http://www.transcriptorium.eu/~htrcontest/contestICFHR2014/public_html/
[10] http://users.iit.demokritos.gr/~kntir/HDIBCO2014/
[11] http://vc.ee.duth.gr/h-kws2014/
[12] http://transcriptorium.eu/~tutorialICFHR/

3.3 Dissemination for content providers

Several talks were given by tranScriptorium partners at different events targeted at content providers:

- P. Schofield. *tranScriptorium.* Digitising Nightingale symposium, organised by the Florence Nightingale Museum and the Howard Gottlieb Archival Research Centre, Boston University.
- J.A. Sánchez. *tranScriptorium.* Dagstuhl Seminar at Leibniz Center for Informatics: "Digital Palaeography: New Machines and Old Texts", 21-24 July 2014, Germany[13].
- P. Schofield, T. Causer and K. Grint. *Many hands make light work. Many hands together make merry work: Transcribe Bentham and crowdsourcing manuscript collections*, International Society for Utilitarian Studies conference, Yokohama National University, 21 August 2014.
- P. Schofield, T. Causer and K. Grint. tranScriptorium *and Transcribe Bentham: How to succeed with scholarly crowdsourcing*, Tokyo University, 25 August 2014.
- P. Schofield, T. Causer and K. Grint. tranScriptorium *and Transcribe Bentham: How to succeed with scholarly crowdsourcing*, Hitotsubashi University, 26 August 2014.
- J.A. Sánchez and E. Vidal. *Interactive Handwritten Text Recognition and Indexing Historical Documents: The tranScriptorium Project* in the *Curious Images* workshop held at the British Library, 18 December 2014[14]

Several public and private institutions made contact directly with the tranScriptorium project members asking about the possibility of future collaborations. Collaboration with the two named centres/institutions in the following list have generated some results, while conversations are ongoing with the other centres/institutions:

1. "IMPACT Centre of Competence"[15].
   tranScriptorium contributed to the DATeCH conference with several papers organized by the IMPACT Centre of Competence.
2. "Biblioteca Nacional de España" (BNE) contacted with UPVLC partner related to tranScriptorium.
   As a result of this contact, an agreement was signed between BNE and UPVLC for transcribing a Spanish collection. The "Plantas" dataset was transcribed in the frameword of this agreement in which also Universidad Complutense de Madrid participated.
3. "Folger Shakespeare Library"[16] from Washington contacted UPVLC asking for information about tranScriptorium.
   UPVLC discussed a possible collaboration with the institution throughout 2014.
4. Stanford University Library[17] made contact at the DATeCH workshop.
5. Real Academia Española de la Lengua[18] made contact at the DATeCH workshop.

[13] http://drops.dagstuhl.de/opus/volltexte/2014/4793/pdf/dagrep_v004_i007_p112_s14302.pdf
[14] https://www.eventbrite.co.uk/e/curious-images-tickets-14438270255
[15] http://www.digitisation.eu/
[16] http://www.folger.edu/
[17] http://library.stanford.edu/
[18] http://www.rae.es/

6. Digital Archives of the National Archives of Finland[19] made contact by email.
7. Instituto Andaluz de Patrimonio Histórico[20] made contact by email.
8. Iesu Communio[21] made contact by email.
9. British Oceanographic Data Centre[22] made contact by email.

3.4 Contacts and dissemination at industrial companies

Some private companies made contact with tranScriptorium partners. The following list shows the list of companies which asked for a face-to-face meeting, and the tranScriptorium project members who gave a presentation:

- G. Mühlberger and V. Romero. XEROX. September 2014.
- E. Vidal and J.A. Sánchez. Contentra Technologies[23] contacted at the DATeCH workshop. May 2014.
- E. Vidal and J.A. Sánchez. I2S[24] contacted at the DATeCH workshop. May 2014.
- E. Vidal and J.A. Sánchez. libnova[25] contacted at the DATeCH workshop. May 2014.
- E. Vidal and J.A. Sánchez. DIGIBIS[26] contacted at the DATeCH workshop. May 2014.
- P. Schofield and T. Causer. Cengage Learning[27]. December 2014.

3.5 In the media

- *Zusammenarbeit Stadtarchiv Bozen und Universitt Innsbruck im Rahmen des EU Forschungsprojekts tranScriptorium fixiert.*

4 Dissemination plan for the forthcoming period

For the third reporting period, the dissemination and demonstration of the project results will continue through publications (scientific papers, presentations at conferences, fairs, or larger exhibitions) in order to disseminate the knowledge acquired during the tranScriptorium project. During the third reporting period of the project, thanks to the expected maturity of the tranScriptorium project and the acquisition of concrete results, we foresee a stronger cooperation in terms of joint publications among the partners. The consortium partners will also go on identifing particular companies, researchers and, in particular, content providers whose activities and interests converge with those of the project.

Furthermore, the most important dissemination instrument of the tranScriptorium project, namely the public website will be regularly updated with news and documents. In particular:

[19] http://digi.narc.fi/digi/?lang=en_US
[20] http://www.iaph.es/
[21] http://www.iesucommunio.com/
[22] https://www.bodc.ac.uk/
[23] http://www.contentratechnologies.com/
[24] http://www.i2s-digibook.com/
[25] http://www.libnova.com/
[26] http://www.digibis.com/
[27] http://www.cengage.co.uk/

- The public deliverables (scheduled at the first and second years) will be available to the general public.
- General presentations and materials will also be regularly published on the public website.
- The list of prototypes will be enlarged in short on the tranScriptorium web page.

Some specific dissemination activities are already scheduled for 2015:

- UPVLC has had accepted a contest on HTR at the *International Conference on Document Analysis and Recogntion (ICDAR 2015)*[28]. The data used for the contest will be a subset of the Benthan data currently used in the project.
- UPVLC has had accepted a contest on handwritten text indexing at the *International Conference on Documente Analysis and Recogntion (ICDAR 2015)*[29]. The data used for the contest will be a subset of the Bentham data currently used in the project.
- UPVLC and UIBK will be participating in the Digital Humanities Conference for the German speaking countries: "DHd-Tagung 2015. Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation" Graz, Austria, 23-27 Feb 2015.
- UCL (Tim Causer) will lead a *Crowdsourcing for Humanities Researchers* workshop, at which he will describe the practicalities of creating a humanities crowdsourcing project and demonstrate tranScriptorium technology at the University of Cambridge, 20 January 2015[30].

5 Exploitation plan for the forthcoming period

One important output of the tranScriptorium project will be the awareness that making easier the access of historical documents can be only tackled by the joint collaboration of different communities, like Document Image Analysis researchers, HTR researchers, content providers, and end users.

The tranScriptorium partners have agreed to an exploitation plan involving the promotion of the technology to widen access to historical handwritten documents through a Virtual Research Environment, as described in the following call "e-Infrastructures for virtual research environments (VRE) EINFRA-9-2015"[31]. This scheme is based on the following ideas:

- Research in Pattern Recognition, Computer Vision, Document Image Analysis, but also in Digital Humanities, Archival Research and related fields has seen unprecedented progress in recent years. Extremely powerful machine learning algorithms, feature extraction methods and document layout analysis algorithm were invented and applied to the HTR field.
- A few years ago archives had digitised just a few precious manuscripts from the middle ages, or personal papers writings by famous artists, politicians or scientists. But today more and more archives understand digitisation as being a natural part of their core mission and are investing significant resources into large scale digitisation endeavours, taking the step from projects to programmes. The next step, beyond digitisiation, is to acquire transcriptions fo these manuscripts images.

[28] http://www.transcriptorium.eu/~htrcontest/
[29] http://www.transcriptorium.eu/~icdar15kws/
[30] http://www.crassh.cam.ac.uk/events/26022
[31] https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2144-einfra-9-2015.html

- We see a growing interest – not only on the part of humanities scholars, but also in the part of the general public – in accessing and transcribing archival documents. Thousands of volunteers are contributing to improvemed accessibility of digitised archival collections through crowdsourcing, and archives are increasingly seeing the benefit of including the general public in creating and improving digital cultural resources.

tranScriptorium partners will promote a Virtual Research Environment that will be submitted in the following H2020-EINFRA-2015-1 call. In addition to tranScriptorium partners, another 8 partners are involved in this proposal.

6 Conclusion

To summarise, our dissemination activities during this second year of the project have largely been about consolidating and generating interest in the technologies being developed. Two main groups of stakeholders have been targeted. First, the scientific community, to generate interest in the innovative techniques being developed; and second, industry, with a view to paving the way for future exploitation. Actions have also been directed at the general public.

In the coming year, activity will continue in all of these areas, with a particular focus on attracting and gaining feedback from end users and content provider users.