

# *tranScriptorium*

---

## D7.1: First Report on Dissemination: Report on Task 7.1 (Year 1)

---

UPVLC, UIBK, NCSR, UCL, INL and ULCC

Distribution: Public

---

**tranScriptorium**

ICT Project 600707 Deliverable D7.1

December 31, 2013



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



Project ref no.	ICT-600707
Project acronym	<b>tranScriptorium</b>
Project full title	<b>tranScriptorium</b>
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, 2013
Actual date of delivery	December 31, 2013
Date of last update	January 20, 2014
Deliverable number	D7.1
Deliverable title	First Report on Dissemination: Report on Task 7.1 (Year 1)
Type	Report
Status & version	Final
Number of pages	10
Contributing WP(s)	7
WP / Task responsible	UPVLC
Other contributors	UIBK, NCSR, UCL, INL and ULCC
Internal reviewer	Joan Andreu Sánchez
Author(s)	UPVLC, UIBK, NCSR, UCL, INL and ULCC
EC project officer	Jose Maria 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@dsic.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/>

© 2013, 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 first reporting period of the project workflow. The document also proposes a dissemination plan for the forthcoming period.

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Dissemination strategy</b>	<b>4</b>
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
<b>3</b>	<b>Dissemination Activities</b>	<b>6</b>
3.1	Dissemination material . . . . .	7
3.1.1	Project documentation . . . . .	7
3.1.2	Project website and other online material . . . . .	7
3.1.3	Demo . . . . .	8
3.2	Publication of research results . . . . .	8
3.3	Dissemination for content providers . . . . .	9
3.4	In the media . . . . .	10
<b>4</b>	<b>Dissemination plan for the forthcoming period</b>	<b>10</b>
<b>5</b>	<b>Conclusion</b>	<b>10</b>

## 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. In this deliverable we provide an overview of all dissemination activities carried out to this effect during the first year of the project.

The first section will be dedicated to identifying the project's stakeholders in more detail. Then follows a record of specific dissemination actions carried out during the M1-M12 period, from the launch of the project website to attendance at various events. Finally we will propose a dissemination plan for the forthcoming period.

## 2 Dissemination strategy

Dissemination activities are directed at two main target audiences, the scientific community and the content providers community, outlined below. A third group, the general public, has also been identified for dissemination throughout the project. In this section we will identify the stakeholders within these categories more precisely, indicating the main ways in which we intend to reach out to them.

### 2.1 Main target audiences

#### 2.1.1 Scientific community

Within the scientific community, results from the TRANSCRIPTORIUM project will be of interest to other research groups in the fields of Crowdsourcing, Computational Linguistics, Handwritten Text Recognition, Image Processing, Document Image Analysis and Pattern Recognition.

An initial list of events being considered in the areas of the project is as follows:

- **ICDAR:** *International Conference on Document Analysis and Recognition.*  
ICDAR is an outstanding international forum for researchers and practitioners at all levels of experience for identifying, encouraging and exchanging ideas on the state-of-the-art in document analysis, understanding, retrieval, and performance evaluation, including various forms of multimedia documents.
- **DocEng:** *ACM Symposium on Document Engineering.*  
The ACM Symposium on Document Engineering provides an annual international forum for presentations and discussions on principles, tools and processes that improve our ability to create, manage and maintain documents. It is sponsored by ACM by means of the ACM SIGWEB Special Interest Group.
- **IbPRIA:** *Iberian Conference on Pattern Recognition and Image Analysis.*  
IbPRIA is a single track conference consisting of high quality, previously unpublished papers, presented either orally or as a poster, intended to act as a forum for research groups, engineers and practitioners to present recent results, algorithmic improvements and promising future directions in pattern recognition and image analysis.
- **ICFHR:** *International Conference on Frontiers in Handwriting Recognition.*  
The ICFHR provides a forum for researchers in the areas of on-line and off-line handwriting recognition, pen-based interface systems, form processing, handwritten-based digital libraries, and web document access and retrieval.

- **DAS:** *IAPR International Workshop on Document Analysis Systems.*  
DAS is a workshop for researchers in the areas of document image analysis, digital libraries, and historical document analysis.
- **ICPR:** *International Conference on Pattern Recognition.*  
ICPR is an international forum for discussions on recent advances in the fields of Pattern Recognition; Machine Learning and Computer Vision; and on applications of these technologies in various fields.
- **CVPR:** *IEEE Conference on Computer Vision and Pattern Recognition.*  
CVPR is one of the most relevant conferences for discussion on recent advances in problems related to image processing and pattern recognition.
- **CAIP:** *International Conference on Computer Analysis of Images and Patterns.*  
CAIP is devoted to all aspects of Computer Vision, Image Analysis and Processing, Pattern recognition and related fields.
- **ICIP:** *IEEE International Conference on Image Processing.*  
The International Conference on Image Processing (ICIP), sponsored by the IEEE Signal Processing Society, is the premier forum for the presentation of technological advances and research results in the fields of theoretical, experimental, and applied image and video processing.
- **ACL:** *Annual Meeting of the Association for Computational Linguistics.*  
The ACL is a leading international scientific and professional society for people working on problems involving natural language and computation. The 50th edition of its popular annual meeting was held in 2012.
- **COLING:** *International Conference on Computational Linguistics.*  
Organised by the International Committee on Computational Linguistics, this biennial conference covers a broad spectrum of technical areas related to natural language and computation, and has a strong focus on networking.
- **TPDL:** *International Conference on Theory and Practice of Digital Libraries.* This conference (*European Conference on Digital Libraries -ECDL-* in the past) is a relevant conference for discussing contents related to Digital Libraries.

### 2.1.2 Content providers community

In this section we consider dissemination from the point of view of paving the way for the future exploitation of TRANSCRIPTORIUM tools, for which we must again identify potential end users or applications.

The most clear application of the TRANSCRIPTORIUM project within the content providers community is the 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 for 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.

Another interesting community are developers of research infrastructure portals (in which the HTR technology shall 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 are being contacted by TRANSCRIPTORIUM, are:

- **Europeana.** Multi-lingual online collection of millions of digitised items from European museums, libraries, archives and multi-media collections.
- **LIBER.** Association of European Research Libraries.
- **APEnet.** Archive Portal Europe.
- **Monasterium.** Europes archives document its political, economic, and cultural development from back to the Middle Ages. Monasterium.net collects many of these historical documents as digital copies for free access.
- **CLARIN.** Common Language Resources and Technology Infrastructure.
- **DARIAH.** The Digital Research Infrastructure for the Arts and Humanities.
- **CENDARI.** Collaborative European Digital/Archival Infrastructure.
- **DASHISH.** Data Service Infrastructure for the Social Sciences and Humanities.
- **EHRI.** European Holocaust Research Infrastructure
- **NeDIMAH.** Network for Digital Methods in the Arts and Humanities.
- **IMPACT** Centre of Competence

### 2.1.3 Society

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

- **Website.** A dedicated project site which serves as a central repository for all reports and documentation on TRANSCRIPTORIUM.
- Publication of news in the media.

## 2.2 Internal dissemination

Internal dissemination is also 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 are being held twice every year for this purpose, hosted in turn by each partner. Meanwhile, day-to-day communication is achieved primarily using online communication channels, such as email and skype. The project website and the wiki are also being 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 during the M1-M12 period for the TRANSCRIPTORIUM project.

### 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 including a project factsheet<sup>1</sup> and a project presentation.

Other material such as project documents and presentation templates were also produced for use by TRANSCRIPTORIUM members.

#### 3.1.2 Project website and other online material

The project website (<http://transcriptorium.eu>) is our main dissemination tool that is available for the general public and is easily accessible. The website presents general project information, the consortium description, scientific publications done during the project, news about the project and public deliverables.

The accesses to the website have been monitored along the year. Table 1 reports the total number of visits for the first year of project.

Table 1: TRANSCRIPTORIUM website access statistics

<b>Number of visits (total)</b>	2,359
<b>Number of unique visitors</b>	1,409
<b>Number of page views</b>	26,051
<b>Average time on site (sec)</b>	226

Table 2 shows the distribution of originating countries. It can clearly be seen that TRANSCRIPTORIUM website has attracted a significant number of visits originating from countries far beyond the original project consortium.

Table 2: TRANSCRIPTORIUM website access statistics per country

<b>Country</b>	<b>Number of visits</b>	<b>% of visits</b>
Spain	913	38,70
United States	297	12,59
United Kingdom	239	10,13
Netherlands	145	6,15
Germany	127	5,38
Greece	126	5,34
Austria	71	3,01
France	47	1,99
Italy	30	1,27
Mexico	30	1,27

TRANSCRIPTORIUM also has profiles on the main online networks, namely, Facebook (<https://www.facebook.com/transcriptorium>) and Twitter (<https://twitter.com/transcriptorium>).

<sup>1</sup>[http://transcriptorium.eu/wp-content/uploads/2012/11/transcriptorium\\_factsheet1.pdf](http://transcriptorium.eu/wp-content/uploads/2012/11/transcriptorium_factsheet1.pdf)

### 3.1.3 Demo

An initial demo of the TRANSCRIPTORIUM functionality has been produced and is being available on the project website.

## 3.2 Publication of research results

At this point in the project, where the emphasis is on getting known, presenting early results, this is arguably the most effective channel of dissemination. It is therefore where most of our efforts and resources have been allocated in the first year.

The following lists give an overview of all papers and communications done by TRANSCRIPTORIUM partners in 2013 (some accepted contributions for 2014 are also included).

### Conferences and Workshops

- F. Álvaro, F. Cruz, J. A. Sánchez, O. R. Terrades, and J. Benedí. *Page segmentation of structured documents using 2d stochastic context-free grammars*, in Proceedings of the the 6th Iberian Conference on Pattern Recognition and Image Analysis (IbPria), 2013.
- D. Martín-Albo, V. Romero, and E. Vidal. *An experimental study of pruning techniques in handwritten text recognition systems*, in Proceedings of the the 6th Iberian Conference on Pattern Recognition and Image Analysis (IbPria), 2013.
- D. Martín-Albo, V. Romero, and E. Vidal. *Interactive off-line handwritten text transcription using on-line handwritten text as feedback*, in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR), 2013, pp. 1312-1316.
- A. Papandreou, and B. Gatos. *A Coarse-to-Fine Skew Estimation Technique for Handwritten Words*, in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR), 2013.
- V. Romero and J.A. Sánchez. *Category-based language models for handwriting recognition of marriage license books*, in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR), 2013, pp. 788-792.
- J. A. Sánchez, G. Mühlberger, B. Gatos, P. Schofield, K. Depuydt, R. Davis, E. Vidal, and J. de Does. *TransScriptorium: a European project on handwritten text recognition*, in 13th ACM Symposium on Document Engineering (DocEng), 2013, pp. 227-228.
- A.H. Toselli and E. Vidal. *Fast HMM-Filler approach for KeyWord Spotting in Handwritten Documents*, in Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR), 2013, pp. 501-505.
- B. Gatos, G. Louloudis, T. Causer, K. Grint, V. Romero, J.A. Sánchez, A.H. Toselli and E. Vidal. *Ground-Truth production in the transScriptorium project*, in 11th IAPR International Workshop on Document Analysis Systems (DAS), 2014, (*accepted*).
- 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, (*accepted*).



### Publications in journals

- 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).

**Given talks.** During the first year, TRANSCRIPTORIUM partners gave several talks in organisations external to the project:

- *Transcribe Bentham*, Twenty-First Century Curatorship series, British Library, London, 28 February 2013
- "*If we can crowdsource Bentham, can we crowdsource anything?*" *The impacts of Transcribe Bentham and collaborative transcription*, Digital Impacts: Crowdsourcing the Arts and Humanities conference, Oxford Internet Institute, University of Oxford, 9 April 2013<sup>2</sup>.
- *Transcribe Bentham: Crowdsourced Transcription of a Scholarly Edition*, COST Easy Tools for Difficult Texts workshop, Huygens Institute, The Hague, 18-10 April 2013<sup>3</sup>.
- *Interactive Transcription and Key Word Spotting in Historical Documents: the transcriptorium project*, DGA workshop, Multimedia Information Processing, Télécom ParisTech, Paris, 1-3 July, 2013<sup>4</sup>.

### 3.3 Dissemination for content providers

An important dissemination activity was the presentation of the project by UIBK representatives at the "Archivschule Marburg" in Germany as a satellite event of the "18th Archivwissenschaftliches Kolloquium" (November 2013) with the title: "Digitalisierung im Archiv. Neue Wege der Bereitstellung des Archivguts (Digitisation in the Archive. New Ways for Providing Access to Archival Material)." This conference with around 180 attendees is one of the most important conferences among German archives. UIBK was invited to provide a paper about the TRANSCRIPTORIUM project for the conference proceedings.

In addition several public and private institutions contacted directly with the TRANSCRIPTORIUM responsables asking for information about the TRANSCRIPTORIUM goals:

- "IMPACT Centre of Competence"<sup>5</sup> contacted with UPVLC asking for information about TRANSCRIPTORIUM. Possible collaboration are currently under discussion.
- "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.
- Responsibles of the library of "University of Connecticut"<sup>6</sup> contacted with UPVLC partner related to TRANSCRIPTORIUM asking for possible collaborations.
- "Folger Shakespeare Library"<sup>7</sup> from Whashington contacted with UPVLC asking for information about TRANSCRIPTORIUM. Possible collaboration are currently under discussion.

<sup>2</sup><http://www.oii.ox.ac.uk/events/?id=573>

<sup>3</sup><http://easytools.huygens.knaw.nl/>

<sup>4</sup>[http://grrhn.insa-rouen.fr/wp-content/uploads/2013/05/invitation\\_seminaire\\_TIM\\_2013.pdf](http://grrhn.insa-rouen.fr/wp-content/uploads/2013/05/invitation_seminaire_TIM_2013.pdf)

<sup>5</sup><http://www.digitisation.eu/>

<sup>6</sup><http://classguides.lib.uconn.edu/content.php?pid=299609&sid=2965500>

<sup>7</sup>[www.folger.edu/](http://www.folger.edu/)

### 3.4 In the media

So far, activities in this area correspond predominantly to online press releases, and to societal dissemination.

For their part, UPVLC contacted the university's Scientific Communication and Innovation department and held an interview with the regional representative of the Spanish language press agency, Agencia EFE. Following this, non-technical articles presenting the project and the problems it addresses have appeared in the online versions of various regional (Las Provincias, La Verdad) and national Spanish newspapers (ABC), and subsequently reposted on other online news compilers.

## 4 Dissemination plan for the forthcoming period

For the second reporting period, the dissemination and demonstration of the project results will continue through publications (scientific papers, presentations at conferences fairs or larger exhibition) in order to spread the knowledge acquired during the TRANSCRIPTORIUM project. During the second reporting period of the project, thanks to the expected maturity of the TRANSCRIPTORIUM project, it is foreseen a stronger cooperation in terms of joint publications among the partners. The consortium partners will also identify particular companies, researchers and mainly 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:

- The first public deliverables (scheduled at the first year) made their way to the general public.
- General presentations and materials will also be regularly published on the public website.
- A demo will be available in short in the TRANSCRIPTORIUM web page.

Some specific dissemination activities are already scheduled in 2014:

- ICFHR 2014. The NCSR group is the co-organizer of the International Conference on Frontiers in Handwriting Recognition in 2014, to be held in Greece. TRANSCRIPTORIUM will be promoted in this conference with the help of NCSR.
- UPVLC submitted a contest on HTR to the ICFHR 2014. The data used for the contest (if accepted) will be a subset of the Benthan data currently used in the project.

## 5 Conclusion

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

Looking ahead, activity will continue in all of these areas, with a particular focus on attracting and gaining feedback from end users.