1 Publishable Summary

This document reports on the activities conducted in the third semester of the PROMISE project, from September 2012 to August 2013. As stated in Annex I of the GA, PROMISE aims at advancing the experimental evaluation of complex multimedia and multilingual information systems in order to support individuals, commercial entities, and communities who design, develop, employ, and improve such complex systems. The overall goal of PROMISE is to deliver a unified environment collecting data, knowledge, tools, methodologies, and the user community.

During the reporting period two project milestones where expected. One has been achieved on time and the other has been slightly delayed since it has been extended in its reach via community feedback:

- **M.V (month 25)**: the third PROMISE annual conference, namely the CLEF 2012 Conference on Multilingual and Multimodal Information Access Evaluation (deliverable 7.9)
CLEF 2012 is the third CLEF conference continuing the popular CLEF campaigns which have run since 2000 contributing to the systematic evaluation of information access systems, primarily through experimentation on shared tasks. Building on the format first introduced in 2010, CLEF 2012 consists of an independent peer-reviewed conference on a broad range of issues in the fields of multilingual and multimodal information access evaluation, and a set of labs and workshops designed to test different aspects of mono and cross-language Information retrieval systems. CLEF 2012 has been hosted by the “La Sapienza” University of Rome, Italy, on 17-20 September 2012.

14 papers and 3 posters were accepted for the Conference and published by Springer in their Lectures Notes for Computer Science (LNCS) series. Two keynote speakers highlighted important developments in the field of evaluation. Peter Clark, Vulcan Inc., USA, focused on how to move from information retrieval to knowledgeable machines. Tobias Schreck, University of Konstanz, Germany, presented research challenges for visual search and analysis in textual and non-textual documents and their evaluation. The community sessions at CLEF 2012 were organized around the presentation of the activities of the other evaluation initiatives and an “Evaluation Clinic” where participants had the possibility of meeting evaluation experts and discussing with them their evaluation problems.

- **M.VI (month 30)**: the final prototype of the PROMISE evaluation infrastructure allowed use to carry out the PROMISE use cases performed during CLEF 2013, namely “Unlocking Culture) (CHiC 2013 Evaluation Lab) “Search for Innovation”
Moreover, it also embeds features for continuously evaluating running systems and applications, and in particular Europeana (deliverable 3.4), which extended the previously designed architecture, as shown in Figure 3 below.

Figure 3: Extended architecture of the evaluation infrastructure with continuous evaluation functionalities.

Figure 1 summarizes the main activities of PROMISE and the relationships and contributions the deliverables due in this reporting period gave to them (deliverables due in this reporting period are in black, deliverables due in the previous reporting period are in grey; post-its indicate related events organized to achieve the goals):

- **D7.9 (month 25) – Third PROMISE annual conference and proceedings**: the CLEF 2012 conference, held in Rome, Italy, from 17 to 20 September 2012, contributed to advancing methodological aspects of the experimental evaluation, reporting on actual evaluation activities carried out during the year in the laboratories, and planning for the next ones in the workshops.

- **D7.10 (month 27) – Third Annual Public Report**: provides a concise overview of the PROMISE activities in its second year of activity and represents an alternative dissemination channel.
- **D2.4 (month 30) – Use case inventory and final specification of the evaluation task:**
  The use case framework and use cases presented in deliverable D2.2 is tested, and the framework and the use cases are revised in accordance with the validation results. Finally, evaluation tasks are formulated based on each of the use cases and the connections between the features of the use cases and the evaluation tasks are discussed.

- **D3.4 (month 30, delayed month 36) – Report on the outcomes of the continuous evaluation activities:**
  Describes the activities on continuous evaluation carried in the context of task 3.4 of workpackage 4. It presents a substantial overview of the use of the DIRECT infrastructure in outside the familiar CLEF campaigns. In particular, the deliverable reports on the use of DIRECT for evaluating Europeana. In addition, the deliverable reports on a number of innovative case studies in continuous evaluation and the implications of the lessons learned for the DIRECT infrastructure. This deliverable has been delayed by six months, from month 30 to month 36. The deliverable builds on activities that ran until month 36, with deliverables in month 36, and allows us to bring together updates to other deliverables that are the result of comments by our reviewers but would otherwise not have had an appropriate deliverable as a landing place. Importantly, the delay of D3.4 from month 30 to month 36 did not affect the overall PROMISE workplan. On the contrary, the delay allowed us to pull together the outcomes of a broad range methodologically diverse evaluation activities and to assess their implications for the DIRECT infrastructure.

- **D7.11 (month 33) – Technology Transfer Day:**
  It has been held at CeBIT 2013 from March 5 to March 9, at the Deutsche Messe in Hannover, Germany. The event involved 13 invited speakers from various academic and commercial institutions and presented various aspects and perspectives of Multilingual Information Access. The workshop provided a forum to discuss and share experiences in the field of Multilingual Information Access.

- **D7.12 (month 33, delay month 36) – Inventory of evaluation resources and evaluation packages:**
  Reports on the evaluation resources used and produced within the PROMISE project between September 2010 and August 2013. This includes all the data sets used and created during the evaluation tasks conducted in the framework of the Cross-Language Evaluation Forum (CLEF). In addition, this report contains the list of “Evaluation Packages” that have been made available in order to make such test collections reusable for benchmarking purposes. It has been delayed to month 36 to be able to include also all the work concerning the collections and language resources acquired during CLEF 2013, ended last September; this has not impacted at all the project work and the other deliverables.

- **D6.3 (month 36) – Report on the outcomes of the third year evaluation activities:**
  Reports the outcomes of the evaluation activities in the third year of PROMISE. PROMISE organizes experimental evaluation activities for multilingual and multimedia information access systems at an international level and on an annual basis. This report presents the outcomes of the CLEF 2012 conference and labs, with emphases on the three main PROMISE use cases participating (unlocking culture, search for innovation and visual clinical decision support). In addition, an outlook on CLEF 2013 organization is presented.
D6.4 (month 36) – Report on the impact analysis for the CLEF Initiative: develop methodologies that measure the scholarly impact of evaluation campaigns and to apply the established workflow (i) for assessing the scholarly impact of the first ten years of CLEF activities (2000-2009), and (ii) for extending a previous study on the scholarly impact of ImageCLEF. We choose to measure the success of CLEF by the scientific impact of the research they foster, i.e., the publications derived from it and the citations they receive. We perform a study of the publication output and of the citation impact using data from Scopus and Google Scholar. Our bibliometric analysis of the CLEF 2000–2009 Proceedings indicates a significant impact of CLEF, particularly for its well-established Adhoc, ImageCLEF, and Question Answering labs, and for the lab/task overview publications that attract considerable interest.

D4.6 (month 36) – Report on the user-centered evaluation methodology: builds and tests a system to support interactive information retrieval experiments both remotely and in the lab. In doing so, it aimed to create a standardised procedure for the conduct of such tests so that data collected could be compared across studies. The resulting system tested in two evaluation campaigns (CLEF and TREC) resulted in a multi-component system that simplified experimental research set-up, managed the procedure for remote processing of participants, and included a rich data collection of user response and log variables. At the same time, the task created a search user interface workbench to simplify the development and inclusion of a search interface in such experiments for testing components, or indeed also for testing how people might perceive a particular algorithm.

D3.5 (month 36) – Final prototype of the evaluation infrastructure with the distributed evaluation protocol: describes the specification and implementation of the PROMISE evaluation infrastructure. It builds on previous D3.3 and adds the specification for the bibliometrics, expert profiling, and topic identification for scientific contributions. It also describes the mapping of the resources to RDF and linked data as well as the final re-design of the overall conceptual model. Finally, it discusses alternatives for component-based evaluation.

D3.6 (month 36) – Report on the semantic exploitation of experimental data: presents several steps in the direction of semantically annotating and interlinking experimental evaluation datasets and scientific publications, with the goal of enhancing their interpretation, sharing, and reuse. We discuss the underlying evaluation workflow and we propose a data model for those workflow areas that are directly related to Expert Search. A topic-centric approach for expert search is proposed, addressing the extraction of expertise topics and their semantic grounding.

D4.4 (month 36) – Report on operational systems as experimental platforms: presents a set of methodological principles, a list of practical considerations, as well as a small number of legal and ethical concerns that living lab experimenters should address, before discussing an extensive list of case studies. The case studies highlight different aspects of the potential and limitations of living lab experiments for information access evaluation, both algorithmic, organizational and practical.

D4.5 (month 36) – Report on the rank analysis techniques for interactive environments: presents an innovative visual analytics environment, called Visual Analytics Tool for Experimental Evaluation (VATE²), which eases and makes more
effective the experimental evaluation process. In particular, VATE supports and improves two typical phases of the experimental evaluation process, namely performance analysis and failure analysis, and introduces a completely new phase, the what-if analysis. All these activities are crucial for conducting rank analysis and they have been carried out by producing an interactive environment.

- **D5.4 (month 36) – Revised collaborative user interface prototype with annotation functionalities:** describes the revised version of the Promise Visual Analytics Environment, detailing the annotation functionalities, the validation activities and the addressed issues.

- **D7.13 (month 36) – Second PROMISE Summer School:** describes the activities carried out for fulfilling deliverable 7.13, the second PROMISE summer school. The activity related to the deliverable, originally planned by August 2013, has been anticipated by 6 months to February 2013 and it has been organized as a winter school instead of a summer school. The topic of the winter school was Bridging between information retrieval and databases and it was organized in the mountain town of Bressanone in Italy. The winter school has been a success under several aspects: quality of the scientific program, participation, and overall satisfaction of the participants.

- **D2.5 (month 36) – Technology transfer report:** gives an overview of how the various activities of PROMISE have contributed to spreading the results and insights gained in during the project lifetime to organisations outside PROMISE itself.

- **D1.8 (month 36) – Report on Legal Continuation:** aims at identifying the better legal form in order to implement the PROMISE activities, beyond the end of the project funded by the FP7. The document analyses the different alternatives which would be suitable to implement the legal entity as well as conducts a survey of the possible assets to be included in the legal entity. The work reported here has served as input to discussions both in the PROMISE consortium and in the CLEF Initiative Steering Committee, discussions which led to the decision of establishing the legal entity

The PROMISE partners strived to organize scientific, dissemination, and education events to involve the research community at large:

- **Panel at IRCDL 2013:** A panel on “Evaluating Cultural Heritage Information Access Systems” has been organized at IRCDL 2013, the 9th Italian Research Conference on Digital Libraries, held in Rome, Italy from 31st January to 1st February 2013.

- **CLEF 2013:** The CLEF 2013 Conference on Information Access Evaluation was held at the Universitat Politècnica de València, Spain, on September 23-26, 2013. The CLEF 2013 conference had more than 200 participants from different academic and industrial institutions. Though the majority of participants come from Europe (149), there is also a considerable interest in CLEF around the world, above all from the Americas with 28 participants, Asia with 22 participants, Australia with 3 participants, and Africa with 2 participants.

Several events have been attended where partners of the project have presented PROMISE, its vision, and its goals, as well as many publications concerning PROMISE results have been published.