

# ACHS'16: First International Workshop on Accessing Cultural Heritage at Scale

Paul Clough  
Paula Goodale  
Information School  
University of Sheffield, UK  
Tel: +44 114 222 2664  
p.d.clough@sheffield.ac.uk  
p.goodale@sheffield.ac.uk

Maristella Agosti  
Department of Information  
Engineering  
University of Padua, Italy  
Tel: +39 049 827 7650  
maristella.agosti@unipd.it

Séamus Lawless  
School of Computer Science and  
Statistics  
Trinity College Dublin, Ireland  
Tel: +353 1 896 1000  
seamus.lawless@scss.tcd.ie

## ABSTRACT

The workshop aims to bring together researchers and practitioners to review and discuss ways of providing effective access to large-scale collections of cultural heritage content. The scale, variety and availability of cultural heritage content, combined with the variety of user groups with respect to background knowledge, specialist experience and needs is challenging in the context of existing access methods. In particular, we consider going beyond keyword search in large-scale cultural heritage digital libraries, in support of exploration and discovery. Our purpose for the workshop is to consider the opportunities and challenges presented by new and existing technologies, as well as the needs and experiences of diverse user communities. Our goal is to assess the current state-of-the-art, to identify opportunities and establish future research priorities, informed by the combined knowledge and experience of academics and practitioners.

## Keywords

Information retrieval, information access, exploration, discovery, digital libraries, cultural heritage, big data

## 1. INTRODUCTION

Large-scale digital libraries are becoming ubiquitous within and between cultural heritage (libraries, museums and archives) institutions, with increasing variety of content (media formats, metadata, and level of detail in text description), diverse user communities (from expert to novice) with differing, often poorly defined, and even unanticipated information needs, and widely varying interfaces to support more exploratory information seeking. In these 'big libraries' with challenges presented by both the volume and variety of content, we can also add the further challenges of needing to support a variety of users and uses.

Approaches to understanding and dealing with these challenges are many and varied, with complementary and competing alternatives in evidence, often using emerging and experimental techniques. Complexity arises in data processing for content

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

JCDL '16, June 19-23, 2016, Newark, NJ, USA

ACM 978-1-4503-4229-2/16/06.

<http://dx.doi.org/10.1145/2910896.2926733>

preparation and presentation; interaction design to allow support for enhanced multimodal approaches to discovery and exploration, which go beyond the search box; and in evaluation of these digital libraries at multiple levels, including system, content and user-specific variables, as proposed for example, via the Interaction Triptych Model [1,2].

In this challenging, and constantly changing environment, it is important to review progress to date, and explore the future potential of emerging trends in information access functionality, interaction capabilities, and underlying techniques and technologies employed, in areas such as knowledge organization, content enrichment, visualization, personalization and adaptivity.

## 1.1 Workshop Description

We focus on challenges and opportunities, current and emerging, in exploration and discovery in large-scale digital libraries, with particular reference to cultural heritage. We consider the underlying enabling technologies, as well as interaction functionalities, and user evaluations. Our goal is to identify the needs of providers and their users, assess the current state-of-the-art, and to identify and prioritize areas of future research potential.

## 1.2 Objectives

The workshop is intended as a forum for knowledge sharing, discussion, and networking. Specific objectives are to:

- Identify outstanding information access challenges and understand their impacts on DL providers and users;
- Review latest developments for enhancing information access in CH digital libraries;
- Work towards a view of future research needs;
- Strengthen connections between academics and industry for potential collaborations.

## 1.3 Topics

The workshop is focused on all aspects of supporting information access, exploration and discovery within large-scale digital libraries, especially within the cultural heritage domain. Key topics include:

- Information discovery, exploration and serendipity
- User-centered information access and evaluation
- Multimedia and multilingual information retrieval
- Information needs and information behavior
- Information organization, ontologies
- Entity-centric information access
- Information extraction, content enrichment, text analytics, natural language processing

- Metadata and linked data
- Visualization of information spaces
- User modelling and adaptation
- Personalization and recommendation

## 2. FORMAT

The workshop will take place over a half day, and will include peer-reviewed papers, with authors allocated 15-20 minutes to present their research. In addition, a panel discussion will further explore matters arising during the course of the workshop. Workshop proceedings will be available online via CEUR.

### 2.1 Audience

The workshop will be of interest to researchers and practitioners working in the fields of digital libraries, cultural heritage, interactive information retrieval, content enrichment, evaluation and user studies, and other areas related to information access in large-scale digital libraries.

## 3. ORGANISERS

**Paul Clough** is Professor in Information Retrieval at the Information School, University of Sheffield. He received a B.Eng. (hons) degree in Computer Science from the University of York in 1998 and a Ph.D. from the University of Sheffield in 2002. His research interests mainly revolve around developing technologies to assist people with accessing and managing information. In particular Paul has published work in the areas of multilingual information retrieval, information access to digital cultural heritage, evaluation of information retrieval systems, geo-spatial search, text-based image retrieval, text re-use and search analytics. Paul is co-author of a book on multilingual information retrieval and over 100 peer-reviewed publications. He is head of the Information Retrieval group at Sheffield.

**Paula Goodale** is a researcher in the Information School, University of Sheffield. She holds Masters degrees in Information Management and in Librarianship, and is completing her PhD in Information Science, at the University of Sheffield. From 2004-2009, she was a Senior Lecturer in Information Management at the University of Huddersfield. Her current research focus is on information access in cultural heritage and digital libraries, with particular emphasis on user requirements and evaluation. Paula has worked on several research projects, including the EC-FP7 PATHS project, and the JISC Search25 project. She was co-chair of the 2<sup>nd</sup> SUEDL workshop.

**Séamus Lawless** is an Assistant Professor in the School of Computer Science and Statistics, Trinity College Dublin. Séamus' research has a strong user focus and all of his work aims to improve the experiences of users when interacting with content and information systems. His research interests are in the areas of information retrieval, information management and digital humanities with a particular focus on adaptivity and personalization. The common focus of this research is digital content management and the application of technology to support enhanced, personalized access to knowledge.

**Maristella Agosti** is full professor in computer science at the Department of Information Engineering, University of Padua, Italy. Group leader of the Information Management Systems (IMS) Research Group of the Department. Permanent Member of the Galilean Academy, Class of Mathematical and Natural

Sciences. Her research interests are information access through search engines and digital libraries, user-oriented keyword-based search systems for structured data, annotation of digital contents, evaluation of digital libraries and archives, and user interaction with digital cultural heritage collections. Principal investigator or senior researcher of several European Commission research projects including CULTURA, PROMISE, EuropeanaConnect, SAPIR, DELOS. Chair from 2009 to 2012 of the Steering Committee of TPD L.

## 4. PREVIOUS WORKSHOPS

The ACHS'16 workshop builds upon previous successful workshops on information access issues in digital cultural heritage, by the current organizers and colleagues:

### 4.1 SUEDL Workshops

The workshop on Supporting Users Exploration of Digital Libraries has been run on two previous occasions, collocated at TPD L 2012 and TPD L 2013 respectively. The first half-day workshop comprised four peer-reviewed papers, keynote, demonstration, and panel discussion. Proceedings are published on the workshop web site [3]. The program for the second full day workshop comprised eight peer-reviewed papers and demos, a keynote, and panel discussion. Proceedings are published in a combined volume with papers from other selected TPD L workshops [4].

### 4.2 ENRICH Workshop

The workshop on Exploration, Navigation and Retrieval of Information in Cultural Heritage was collocated at SIGIR 2013. The program comprised six peer reviewed papers, seven posters and demos, keynote and plenary session. Papers are available to download via the workshop web site [5].

## 5. REFERENCES

- [1] Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., and Papatheodorou, C. (2007). Evaluation of digital libraries. *Int. J. Digital Libraries*, 8(1), 21-38.
- [2] Steiner, C.M., Agosti, M., Sweetnam, M.S., Hillemann, E.-C., Orio, N., Ponchia, C., Hampson, C., Munnely, G., Nussbaumer, A., Dietrich, A., and Conlan, O. (2014). Evaluating a digital humanities research environment: the CULTURA approach. *Int. J. Digital Libraries*, 15(1), 53-70.
- [3] Agirre, E., Fernie, F., Otegi, A. and Stevenson, M. (2012). *Proc. 1<sup>st</sup> International Workshop on Supporting Users Exploration of Digital Libraries*, (27 September, 2012, Valletta, Malta). SUEDL Workshop. <http://ixa2.si.ehu.es/suedl/SUEDLproceedings.pdf>.
- [4] Bolikowski, Ł., Casarosa, V., Goodale, P., Houssos, N., Manghi, P., and Schirrwagen, J. (Eds.). (2014). *Theory and Practice of Digital Libraries - TPD L 2013 Selected Workshops: LCPD 2013, SUEDL 2013, DataCur 2013: Revised Selected Papers*, (26 September, 2013, Valletta, Malta), Springer LNCS, Communications in Computer and Information Science, Vol 416. DOI=<http://dx.doi.org/10.1007/978-3-319-08425-1>.
- [5] Workshop on Exploration, Navigation and Retrieval of Information in Cultural Heritage (ENRICH) (1 August, 2013, Dublin, Ireland). <http://www.cultura-strep.eu/events/enrich-2013/programme>