

PATHS is funded by the European Commission FP7 programme under the Digital Libraries and Digital Preservation objective, and brings together six partners to research and develop a system to act as an interactive personalised tour guide through digital library collections. The project's main objectives are to:

- *provide users with innovative ways to access and utilise the contents of digital libraries that enrich their experiences of these resources;*
- *extend the state-of-the-art in user-driven information access by applying language technologies to analyse and enrich online content;*
- *take a user-centred approach to development to accommodate the needs, interests and preferences of different types of users*

Summary of activities

During its first year, the PATHS project has been carrying out research to design and specify the PATHS system. The project has adopted a user-centred approach to system design and has been working to gather information about end-user requirements through on-line surveys, interviews, lab trials, analysis of query logs and analysis of sentiments expressed online about cultural heritage information services. Content has been collected from Europeana and Alinari 24 ORE, analysed and enriched using Natural Language Processing techniques and is now available for integration in PATHS prototypes. Work has also been underway to define the PATHS system architecture and to specify the standards and protocols which will be used to provide a robust yet flexible system. The user interfaces to the system have been designed and a high-fidelity prototype of the design has been created for demonstration and evaluation. Implementation of the PATHS system architecture has started, beginning with the virtual data repository and data integration sub-systems.

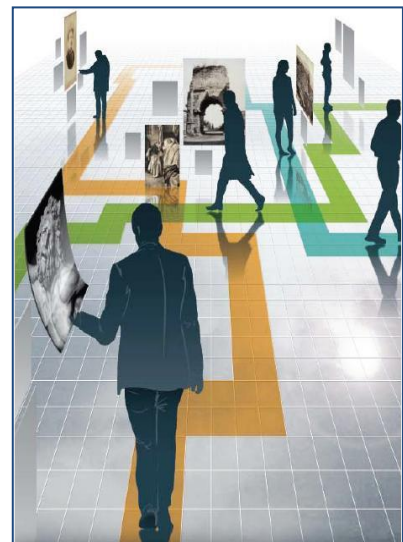
Building good communications amongst the partner network began at the project kick-off meeting in Sheffield and is supported by regular Skype meetings. Partners have been active in disseminating news about the project's activities at conferences and through publications. A team from PATHS took part in a Europeana Hackathon held in September 2011 and won third prize for their entry 'Wikiana'.

The project has now begun work to implement the first PATHS prototype, which will be available for demonstration and evaluation in the second quarter of 2012.

User requirements and functional specifications

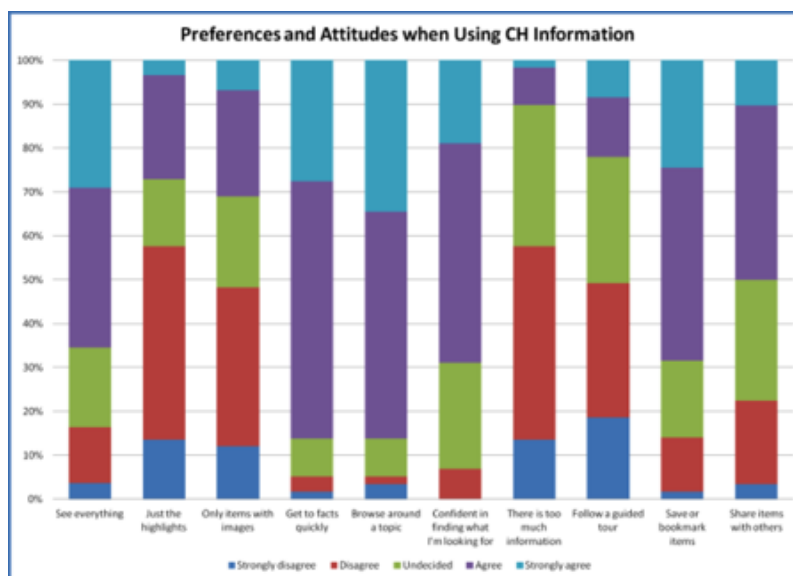
The PATHS project has adopted a user-centred approach to system design, aiming to gather feedback from potential end users at every stage of development. Our development cycle began with user requirements analysis involving desk research, quantitative and qualitative surveys, and controlled experiments.

A review of previous user studies, existing examples of paths and path-creation tools, and literature sources helped inform the requirements gathering. In early summer 2011 an online survey to collect input from potential end-users of PATHS and gain insights into their characteristics, information behaviour and attitudes towards using cultural heritage information. A series of interviews were carried out with expert end-users exploring the concept of paths and helped assess their requirements as producers and users of cultural heritage information.



The results of this research enabled a set of user profiles to be developed. Not surprisingly similarities and differences between user groups were identified. For example, curators, academic researchers and professionals in promotional roles create paths for consumption, while museum educators and teachers are more likely to focus on enabling other users to create paths.

PATHS creation activities were validated in a series of task-based experiments and then used to derive a generalised conceptual model of user interaction with the PATHS system. The key elements of engagement with the PATHS system are: developing a concept or idea for a path; collecting relevant resources; creating a path; communicating with others about a path; and consuming paths. Use cases based on the user profiles and the interaction model helped to derive a comprehensive set of implied user requirements.



The results of the user requirements analysis formed the primary inputs to the functional specification for the first PATHS prototype.

The project completed a review of the State-of-the-art in educational informatics, information retrieval and semantic similarity and relatedness. Surveys of sentiment towards cultural heritage online are also helping inform developments.

Content processing and enrichment

The project has been provided with content by Europeana (www.europeana.eu) relating to collections from the UK and Spain, and content from project partner Alinari 24 ORE. In total, metadata for over 2 million items was received. A PATHS metadata format was defined consisting of an extension to the Europeana Semantic Elements schema with the additional elements allowing for items in the collection to be enriched by PATHS.

After analysis the metadata was processed and enriched by the creation of links to related items in the collection and to external sources. Natural Language Processing techniques were used to identify relevant pieces of information used for generating links including people, organisations, dates, locations and subjects.

The first experiments were made in using ontologies to extend the information available in the Europeana collection. The subject field was automatically compared to some relevant vocabularies – these were the Library of Congress Subject Headings (in English and Spanish) and English Heritage's thesauri. The match suggests that these vocabularies may be used to enable faceted browsing in the PATHS user interface.

Links within collections are being produced to connect each item to the 25 most relevant items using an approach based on Latent Dirichlet Allocation model. In addition, links to articles in Wikipedia are also being generated to provide users with background information about items in the collection and with a richer experience. In future these links may also be used to categorise collections using Wikipedia's category hierarchy.

The results of these processing techniques is the creation of a multi-layered information network which will form the basis by which paths between items can be suggested and used to navigate the collection. An example of a multi-layered network is illustrated here. The content processed during this year will provide the basis for the first prototype of the PATHS system.

System Architecture

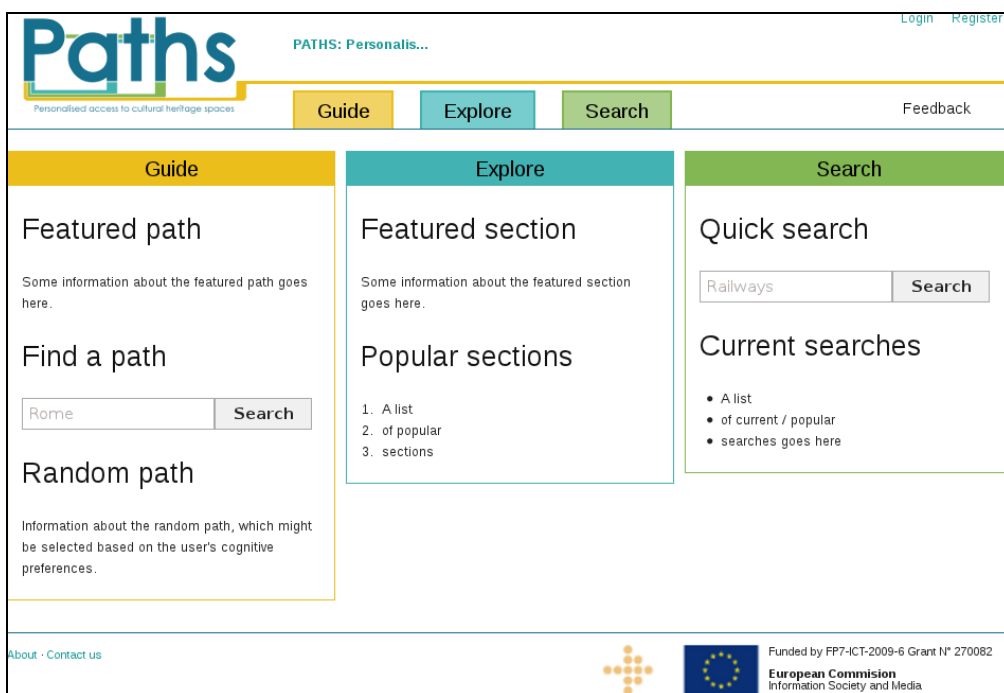
Our objective is to establish a flexible system architecture to allow for the development and integration of PATHS applications with existing collection management systems in use by cultural heritage organisations. The functional and technical requirements for the system in terms of hardware, software, scalability, robustness and security have been analysed. A series of studies and technical trials were carried out to identify the protocols and standards relevant for the system architecture. The user requirements and functional specification are taken into account in the specifications for the system architecture.

The data model and information flow is an important part of the PATHS system architecture. The architecture needs to deal with a vast amount of data while delivering the performance and robustness needed to provide an acceptable user experience. We defined a model for the storage and representation of the content as well as standard interfaces and protocols to enable the content to be used efficiently by the components in the system. As RDF/SPARQL data models are outperformed by RDB/SQL based data models, PATHS is implementing a hybrid model with a hierarchical selection process using traditional RDB/SQL queries thus limiting the number of triples needed to be processed by SPARQL queries. Web Service APIs will be used to communicate between the end-user applications and the underlying system to provide flexibility and modularity of components.

The development team has begun the implementation of the system architecture in preparation for the implementation of the first PATHS prototype in 2012. The virtual data repository has been implemented and work is now underway on the data integration sub-system.

User Interaction and Interface design

The focus of work has been to design the user interfaces for the first PATHS prototype. Following our user-centred methodology, the findings from the user requirements analysis together with the functional specification and accepted design principles provide the foundation for the design process. Hand-drawn storyboards using low-fidelity methods were evaluated and then developed further using high-fidelity methods to provide a fully-functioning high fidelity interaction design in HTML which is based on the content collected and processed by the project. This provides a demonstrable object to show user groups and is enabling heuristic evaluation of the designs.



The high-fidelity user interaction design mock-ups form the basis for the design and implementation of the user interface for the first PATHS prototype in spring 2012.

Work is also underway to gather user transaction logs which will be analysed to identify patterns of behaviour that can be mined to help make recommendations and aid navigation/discovery. So far around 700,000 user sessions have been downloaded from Europeana with over 14,500,000 log entries. Log files from other Cultural Heritage institutions are also being collected to enable a comparison of user behaviour across different collections.

Evaluation and Field Trials

Work has been underway to plan the evaluation methodology for the project in preparation for the evaluation of the first PATHS prototype in the second quarter of 2012. Objective and subjective measures are being defined including measures to assess the accuracy, reliability and scalability of the system, and engagement with users in laboratory and field settings.

The University of Sheffield organised a TREC track on evaluating Information Retrieval systems over user sessions, and this may be used by PATHS in its evaluation trials.

Dissemination

During the first year of the project we have established the project website (<http://www.paths-project.eu/>) and an initial set of dissemination materials including a project brochure, posters and the project newsletter. The partners have been active in disseminating the initial project results taking part in 29 events in Europe, the USA and Australia. Several papers about the project and the research have been published or are in the process of being published. Copies of reports, presentations and pre-prints of conference papers and other resources can be found on the project website <http://www.paths-project.eu/eng/Resources>.



In September 2011 a team from PATHS entered a Europeana hackathon held during the final conference of the Europeana Connect project. The team, with members from the Universities of Sheffield and the Basque Country, developed an application called "Wikiana" which augments Wikipedia articles with relevant items selected from Europeana using the Europeana API. The idea was to give users an overview of the data available in Europeana and to provide a link from the item in the Wikipedia article to the specific item in Europeana. The entry from PATHS received third prize.

PATHS groups have been set up on LinkedIn and Facebook, and Twitter is being used to distribute news to followers and also via a Tweet feed embedded into the project website.

Future work

Work is currently underway to implement the first PATHS prototype which is due for release in April 2012. A series of demonstrations and laboratory based user evaluation trials are planned for the second quarter of 2012, and these will be complemented by testing of the system architecture. The results of the evaluation of the prototype will feed into the specifications of the second PATHS prototype. Work on content collection and processing will continue with further experiments with content enrichment with the aim of developing an augmented data set for implementation in the second PATHS prototype in early 2013. Plans are underway for a PATHS workshop which will be held in 2012.

Contacts

Coordinator: Mark Stevenson, University of Sheffield, M.Stevenson@dcs.shef.ac.uk

Project manager: Kate Fernie, MDR Partners (Consulting) Ltd., kate.fernie@mdrpartners.com

PATHS has 6 partners from 5 countries: University of Sheffield (UK), i-sieve Technologies Ltd (Greece), Asplan Viak Internet (Norway), Alinari 24 ORE (Italy), MDR Partners (Consulting) Limited (UK) and Universidad del País Vasco / Euskal Herriko Unibertsitatea (Spain).