



Grant Agreement No. **ICT-2009-270082**
Project Acronym **PATHS**
Project full title **Personalised Access To cultural Heritage Spaces**

D 8.1

Final report

Authors: Kate Fernie, Jill Griffiths and Rob Davies (MDR)

Contributors: Mark Stevenson, Paul Clough, Samuel Fernando, Nikos Aletras (University of Sheffield)

 Eneko Agirre, Arantxa Otegi, Aitor Soroa (University of the Basque Country)

 Runar Bergheim, AVINET

 George Chrysochoidis, i-sieve

Project funded under FP7-ICT-2009-6 Challenge 4 – “Digital Libraries and Content”	
Status	Final
Distribution level	Public
Date of delivery	
Type	Report
Project website	http://www.paths-project.eu
Project Coordinator	Dr. Mark Stevenson University of Sheffield

Table of Contents

1 Project objectives.....	3
2 Consortium.....	4
3 Project achievements.....	5
4 Target users and their needs.....	6
5 System and prototypes.....	7
5.1 System architecture.....	8
5.2 Data model and database.....	8
5.3 PATHS Web API.....	8
5.4 PATHS web application.....	8
5.5 User interface.....	9
5.6 Recommender system.....	11
5.7 Mobile application.....	12
6 Semantic Enrichment of Cultural Heritage Content.....	13
6.1 Data	13
6.2 Motivation for Semantic Enrichment.....	13
6.3 Semantic Enrichment techniques.....	13
6.3.1 <i>Identification of Key Entities</i>	13
6.3.2 <i>Item Similarity</i>	14
6.3.3 <i>Typed Similarity</i>	14
6.3.4 <i>Background Links</i>	14
6.3.5 <i>Hierarchies</i>	14
6.4 Application of Semantic Enrichment.....	15
6.4.1 <i>Semantic Enrichment Using Open Source Software</i>	15
6.4.2 <i>Representation of Semantic Enrichment using Europeana Data Model</i>	15
6.5 Web service for semantic enrichment.....	15
7 Evaluation and field trials.....	16
7.1 Gathering User Requirements	16
7.2 User evaluation.....	16
7.3 Applications for Paths in Digital Cultural Heritage.....	18
8 Summary of activities.....	19
9 Impact and exploitation.....	21
10 Further information.....	22
11 References.....	23
11.1 Project Deliverables.....	24

1 Project objectives

PATHS set out to research and develop techniques to improve access to very large digital library collections being offered to end-users.

Significant amounts of cultural heritage material have been made available through online digital library portals as a result of digitization initiatives in recent years. Such large collections can be difficult to navigate for many users, especially those who lack specialist knowledge about the content or structure of the collection. The metadata that is available can limit the information retrieval that cultural institutions are able to offer their users. In collections where content is aggregated, such as Europeana (<http://www.europeana.eu>), from many institutions creating metadata using diverse cataloguing systems these issues are compounded. PATHS aimed to demonstrate the potential for improving users' experience of exploration and discovery within cultural heritage collections by exploiting natural language processing to enrich metadata and by implementing state-of-the-art systems in user-driven information access.

Throughout the three years of the project PATHS has researched, developed and evaluated techniques to improve the presentation of digital library content to end users. The team have worked to develop understanding of user profiles, needs and requirements and to bring this user focus into the development cycle. Specific research objectives:

- Developing a system that enables users to navigate cultural heritage collections by using pathways, connections between objects and tools that visualise the topic content of the whole collection.
- Advancing techniques to analyse and enrich the content metadata available for information retrieval and allowing relevant background information to be offered.
- Working with users to identify their requirements and to evaluate the system to inform the development of a user-driven information retrieval system.
- Developing tools to enable users to create pathways and to make connections between content items that are of interest.
- Establishing an exploitation plan for PATHS results

PATHS has worked with Europeana and Alinari 24 ORE in exploring the potential to enrich content metadata to demonstrate the potential to offer users recommendations and contextual information. The natural language processing techniques developed in PATHS have had a significant impact in the user interface by enabling novel browsing and improved information discovery. The knowledge developed through the project will be invaluable in the development of the next generation of portals.

2 Consortium

Six partners formed the project consortium:

The University of Sheffield - the Information Retrieval Group (based in the Information School) and the Natural Language Processing Group (based in the Department of Computer Science) both participated in PATHS. Together these groups form one of the largest language technology and information science clusters in the EU.

i-sieve Technologies Ltd - a spin-off company of the National Centre for Scientific Research “Demokritos” founded in late 2004 in Athens, Greece. i-sieve has developed a suite of tools specifically targeted to online data collection and sentiment analysis. i-sieve serves corporate clients and media firms across Europe with business intelligence applications through online media analysis delivering sentiment snapshots and trend detection insights.

Asplan Viak Internet Ltd - a Norwegian company specialized in development of Internet-based spatial-enabled and database driven end-user applications, established in 2001. The development team consists of skilled and experienced ICT architecture and development professionals with rich experience within the domain of cultural heritage, e-government, e-learning, tourism, planning and environmental management disciplines.

Alinari 24 ORE - Fratelli Alinari Archive merged in the year 2007 with the Sole 24 Ore to create a new company called Alinari 24 ORE, which company dissolved in 2012. The Fratelli Alinari Archive was founded in Florence in 1852 and is the guardian of a photographic ‘corpus’ that includes over 5,500,000 pictures, both historical and contemporary. Alinari 24 ORE was a partner in PATHS until December 2012.

MDR Partners (Consulting) Ltd - a leading specialist partnership working together with libraries, archives, museums, information providers and other organisations in cultural heritage and education sectors. MDR has a strong track records in the creation of new projects, establishing project partnerships and networks, as well as working with European programmes and funding bodies.

The University of the Basque Country - is a teaching and research institution officially founded in 1985. The IXA Research Group, which participates in PATHS, was created with the aim of promoting the modernization of Basque by means of developing advanced computational resources for it. The IXA group has a strong international standing in the areas of Semantic Interpretation, Semantically based Information Retrieval and Information Extraction.

PATHS is coordinated by University of Sheffield with MDR Partners as project manager.

A list of partners can be found at: <http://www.paths-project.eu/eng/About/Project-partners>.

3 Project achievements

One of the major achievements of the Paths project has been to demonstrate the practical benefits and technical feasibility of enriching the metadata for cultural heritage collections as a means of improving content retrieval, supporting innovative discovery and exploitation. This addresses a critical issue for cultural heritage institutions across Europe who hold vast quantities of quality content in digital libraries that are currently never found unless explicitly sought.

Positive feedback from users involved in the evaluation of the PATHS prototypes confirmed that the project achieved its objective of developing tools that add value to digital libraries and enrich user experiences. The interactivity offered by PATHS tools and the potential for expert and non-expert users to use content to create narratives, tell stories and make personal collections has the potential to have real impact on the development of the next generation of cultural heritage services.

PATHS has made a significant contribution to research into the application of semantic enrichment techniques to cultural heritage content demonstrating the potential to enrich the simple content metadata enabling novel browsing and information discovery. The techniques used enriched items with links to similar content, created links to related Wikipedia articles, and enabled the collections to be thematically organized into a semi-automatically created hierarchical structure.

PATHS explored the potential of generating recommendations to recommended items and personalized recommendations based on user profiles, query logs and the similarity of content items. The project contributed to research in this important area.

A robust and holistic evaluation methodology involving demonstrations, laboratory testing, field trials and system testing was established, enabling a comprehensive evaluation of the PATHS system.

The project provided an important test-bed for Europeana content and the Europeana Data Model (EDM) making recommendations on extensions to the EDM schema to manage semantically enriched content.

The knowledge developed through the project with regard to information retrieval, indexing and Web APIs have been invaluable to the development of the next generation of portals. A prototype content enrichment web service was released and has been adopted for development by the LoCloud project. The next generation of Avinet's map portal, ADAPTIVE, is being built exploiting PATHS technology in its search engine and is a first step on the way towards commercial exploitation of Paths products.

The project has established an exploitation plan for each of the products developed through PATHS identifying areas with potential for future research, commercial development and the delivery of consultancy services.

All research activities have lead to scientific publications, and the project results form a core technology with expectations for further exploitation in the Cultural Heritage domain.

4 Target users and their needs

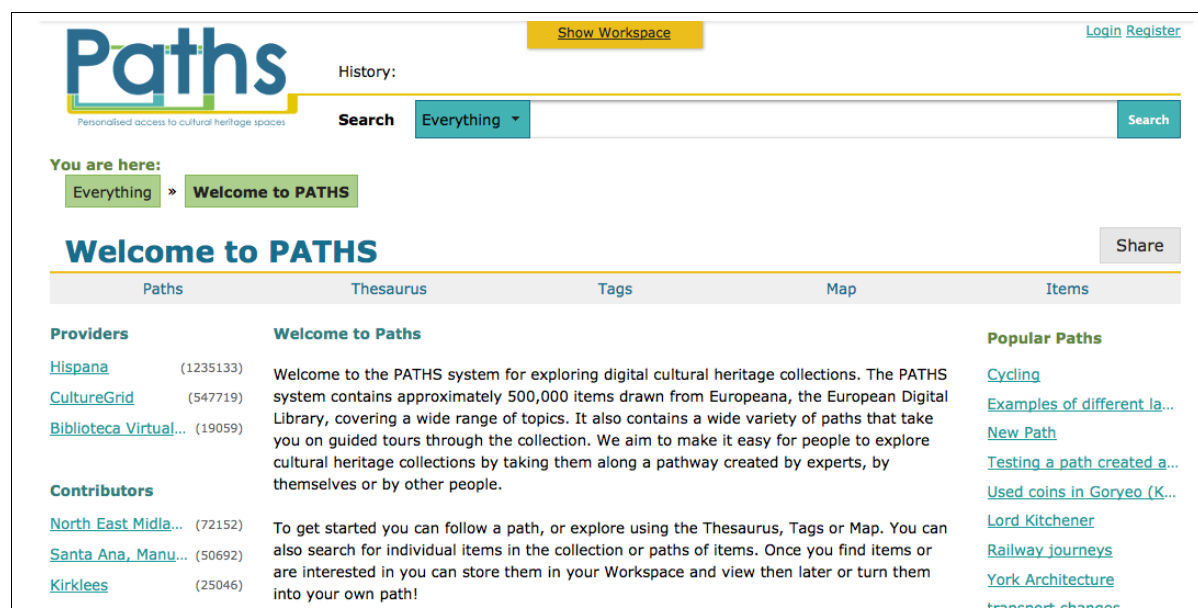
The main target user groups for PATHS are:

- **Cultural heritage institutions** (libraries, museums and archives) that are making increasing quantities of digital content available online and who wish to make their collections more accessible to citizens, improve users experiences of using their systems and to benefit from sustainable techniques for content enrichment.
- **Technology providers** who are developing information retrieval and content management system and who wish to benefit from new techniques in order to offer state-of-the art systems.
- **Scientific community** who are active in the research fields and wish to share result results and to advance new techniques.
- **Europeana** which is seeking to implement content enrichment services and to enable story-telling to improve both the effectiveness of information retrieval and its users ability to interact with its content and service.

Partners in PATHS addressed a number of research challenges:

- Analysing the very simple metadata available for content items in Europeana information retrieval in order to find the best ways of processing and enriching the metadata to improve information retrieval.
- Working in the cloud as a distributed team making best use of open standards and open source software to develop and test the components that supply the functionality of the PATHS system.
- Innovating in the design of the user interfaces to offer new ways for users to explore and navigate their way in very large digital library collections.
- Embracing user-centred design and working with user groups to first identify their requirements and later to evaluate and test the PATHS prototypes.
- Finding ways of offering personalisation and recommendations to users of cultural heritage digital libraries.

5 System and prototypes



PATHS prototype

The PATHS project has developed two prototype systems for testing and evaluation by end users, cultural institutions and others. Our aim has been to demonstrate a collection discovery environment that combines, search functionality with narrative pathways and visualisation techniques. These are based on user requirements and functional specifications gathered early in the project, with user feedback on the first prototype informing the development of the second. The system offers three ways into collections:

- Users can follow Pathways (or narratives) published by other users; pathways can be followed from beginning to end or users can leave the path at any time by following recommendations and links. Users who are logged in to the system can save items to their workspace and then create, edit, publish and share Pathways.
- Users can explore the collections using the Thesaurus, Tags or the Map, which each provide visualisations of the digital library.
- The system search functionality supports discovery of collection items and Paths.

The PATHS prototype incorporates data provided to the project by Europeana for its research. The PATHS demonstrator is available at <http://explorer.paths-project.eu/>.

5.1 System architecture

The PATHS system includes software, hardware and data. The system architecture has three logical layers:

1. A data layer consisting of a data repository sub-system (the PATHS Database and logical data model) and the content processing sub-systems described in section 3 above
2. An application layer with seven sub-systems including data integration; user profile management; Path authoring; recommendations; information retrieval; presentation; and a web service API
3. A client layer – the web application for end users.

The system documentation is available here:

<http://www.paths-project.eu/eng/Prototype/PATHS-system>

5.2 Data model and database

The data model is implemented as a combination of SQL and non-SQL databases and indexes combining the power of the PostgreSQL RDBMS with that of Apache Solr inverted index and Virtuoso RDF triple store. All relationships and references in the data model are implemented using persistent URIs as foreign keys. This allows for flexible integration between the three data stores.

5.3 PATHS Web API

The web service API is implemented on top of a wide range of server components and controls all data I/O operations towards the data layer. It consists of more than 30 different web methods grouped into seven Web Services. The Web Services communicate over the HttpGet, HttpPost, Soap and Soap 1.2 protocols. The default return format is JSON but the services are also capable of delivering XML.

5.4 PATHS web application

The PATHS prototype is implemented on a web server platform, which runs on the Windows 2008 Server operating system. The core web server is Internet Information Server (IIS) integrated with Apache Tomcat Servlet Container (Tomcat) to enable components such as the search server SOLR.

All data in the system are stored in the PATHS PostgreSQL database including item records semantically enriched by the project (see section 3 above) and also data for users, paths and nodes created within the user interface. The SOLR search server is used to index all items, nodes and paths. A Virtuoso Triple Store is used to support the recommender service.

The end-user interface is a Python client application that calls on the PATHS Web Service API for all of its data operations.

5.5 User interface

One of the objectives of PATHS has been to take a user centred approach in designing the user interface to the system. Our aim has been to focus on the goals that users wish to

achieve delivering the functionalities, while supporting both good interaction design and innovative means of exploring the digital library content.


Our approach has been to involve users at all stages of the research and development cycle. We began the cycle by gathering user requirements and using these to prepare functional specifications for the system. Next storyboards of the user interface were prepared to gather feedback. Development of high-fidelity interaction design allowed further feedback to be gathered from users before the user interface was implemented in the first system prototypes. User evaluation and testing of the prototype then helped to inform the next development cycle and the development of a second system prototype.

Interface components

The core components of the user interface include:

- **Search** – a free text search box allows users to enter a search term and use it to search the items and paths in the digital library collection; facets are provided to allow users to narrow down their searches.
- **Exploration** – a thesaurus allows users to explore the digital library collection using the concepts within the thesaurus hierarchy; a tag cloud based on the thesaurus allows users to visualise the frequency of concepts within the digital library collection; a topic map provides an alternative way of visualising the information space defined by the concepts in the thesaurus and their frequency.
- **Item display** – displays information about the item including both the item record and related information identified by semantic enrichment of the content and the PATHS recommender service.
- **Path following** – this component allows users to follow an existing path either by beginning at the first node, by jumping to nodes discovered in searches or by using the path overview.
- **Path creating** – a work space is provided to allow users to collect items that they have discovered in searching or exploring the collection; while the path creation component allows users to move items from their workspace and arrange them in a path adding their own descriptions, text-only nodes and creating branches in paths as needed.
- **User** – this component allows users to register and to create an account and profile in the system.

The thesaurus implemented in the second PATHS prototype was compiled using natural language processing techniques and topic hierarchy derived from Wikipedia categories.



Personalised access to cultural heritage spaces


[Show Workspace](#)
[Login](#)

History: [Everything \(items\)](#) [Personalised Acces...](#) [Everything \(thesa...](#) [Design Council \(th...](#)

Search

You are here: [Everythi...](#) » [Design Coun...](#) » [Britain Can Make...](#)

Britain Can Make It

Paths	Thesaurus	Tags	Map	Items
Providers CultureGrid (1080)	Everything Alloy Chancel Coin Copper Crafts Study Centre Denarius Design Council Abram Games Britain Can Make It Loggia Royal Institute of British Architects University of Brighton Victoria and Albert Museum Charles Rennie Mackintosh Design	725 topics 8 topics 5 topics 136 topics 46 topics 11 topics 8 topics 31 topics 0 topics 4 topics 0 topics 0 topics 0 topics 0 topics 0 topics 0 topics 0 topics	547780 items 507 items 407 items 28089 items 9505 items 1267 items 361 items 3452 items 68 items 1080 items 20 items 20 items 335 items 22 items 22 items 20 items	0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths 0 paths
Categories Item (1080) Physical object (1080)				Selected Items 
Contributors No contributors found				Help <p>Nulla posuere tincidunt vestibulum. Donec imperdiet aliquam urna, ut ultricies dui bibendum id. Curabitur at ultricies orci.</p>

Second PATHS prototype: discovery by thesaurus

The topic map enables users to visualise the topics in the entire digital library collection and to zoom in to explore topics which are of particular interest.

[Login](#) [Register](#)

Paths

Personalised access to cultural heritage spaces

History: Personalised Access... Everything (thesaurus) Everything (map)

Search

English ▼

Search

You are here:

Everything

 >

English

English

Share

Add to Workspace

Paths
Thesaurus
Tags
Map
Items

Providers

[CultureGrid](#) (547309)

Categories

[Item](#) (547309)

[Image](#) (380616)

[Physical Object](#) (93077)

[Text](#) (53933)

[Path](#) (43)

Contributors


[North East Midlands](#) (72096)

[Kirklees](#) (25018)

[MSC Scheme](#) (547)

[North of England Museum](#) (526)

[Atkinson, Peter](#) (414)



The map displays a network of interconnected topics. Central nodes include 'Society', 'Nature', 'Health', 'Law', 'Science', 'Mathematics', 'Computers', 'Technology', 'Arts', 'People', and 'Home'. Other visible nodes include 'Social sciences', 'Communities', 'Organizations', 'Matters', 'Natural sciences', 'Human names', 'Design', 'Crafts', 'Tort law', 'Sausage', 'Belgium', 'Belgian scientists', 'Ribblesdale', 'ts by country', 'Robert Clavering', 'Puffin', 'Oligarchs', 'Robin Hood', 'Defence', 'Matter', 'Social issues', 'Physical Object', 'Image', 'Text', 'Path', 'Providers', 'Categories', 'Contributors', 'North East Midlands', 'Kirklees', 'MSC Scheme', 'North of England Museum', 'Atkinson, Peter', 'About Contact Cookies', 'Powered by Leaflet - Map Data by the PATHS project'.

Selected Items

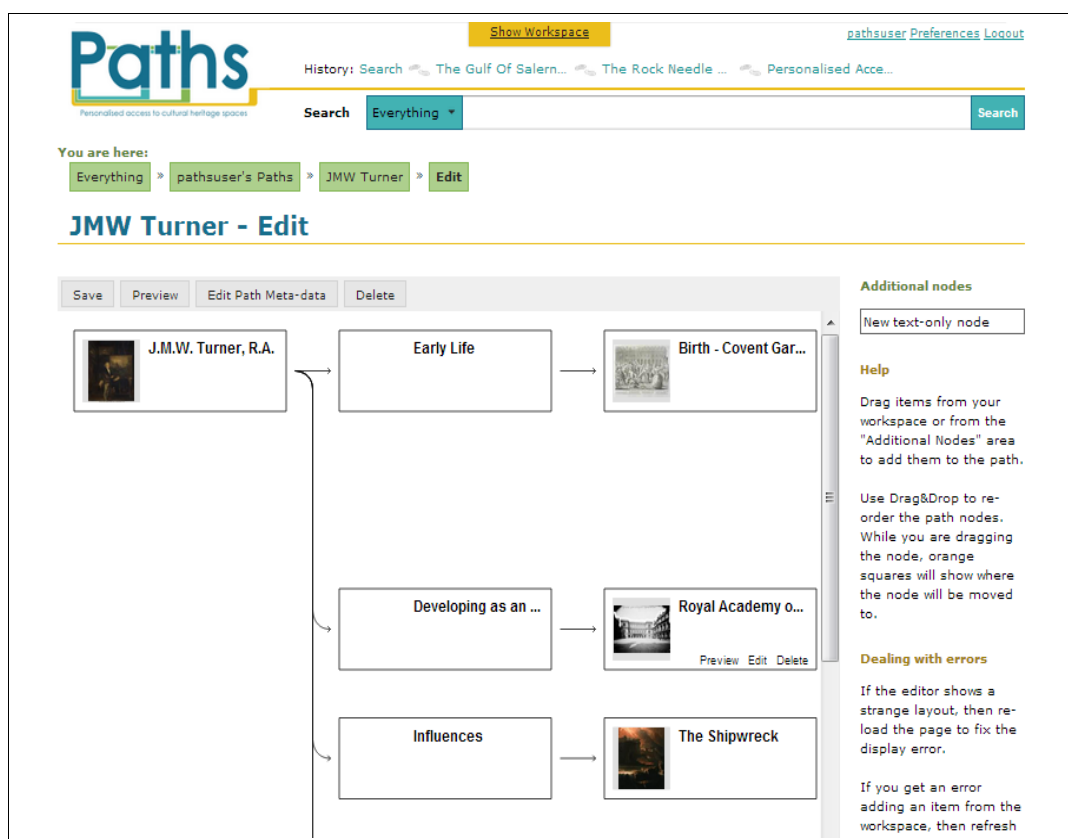
Help

The map shows the topic space of the Paths system. Explore it like you would any other web-map.

If you zoom in, then in addition to the topics, you will also be shown individual items.

Second PATHS prototype: discovery by topic map

The path creation interface enables users to drag-and-drop items from their workspace to the editing space where they can construct branching paths. Text nodes can be added to enable users to add their own descriptive content.



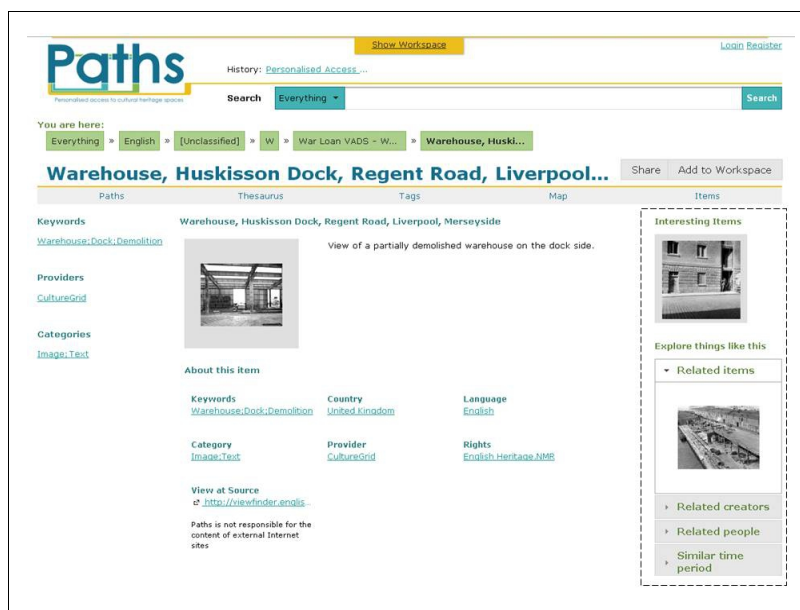
Second PATHS prototype: Path creation

5.6 Recommender system

The recommender system that is currently implemented in the PATHS prototype (illustrated below) offers users non-personalised recommendations of “interesting items” related to the item of content they are currently looking at. The aim of this functionality is to improve horizontal exploration of the collection.

For the “interesting items” functionality, the similarity between each pair of items is computed using a state of the art approach based on Latent Dirichlet Allocation over the text, allowing users to quickly find related items when browsing. An evaluation dataset was crowd-sourced to enable us to assess this approach [Aletras, et al, 2013].

The project team also carried out research into ways of providing personalised recommendations. This included analysis of Europeana query logs to see if there was enough data to implement “*people who viewed this item also viewed*” recommendations. In addition a number of recommender system algorithms were evaluated; this research concluded that combining simple-keywords with query logs and similarity measures produces the best results. These findings will be used in future to improve the recommender system.



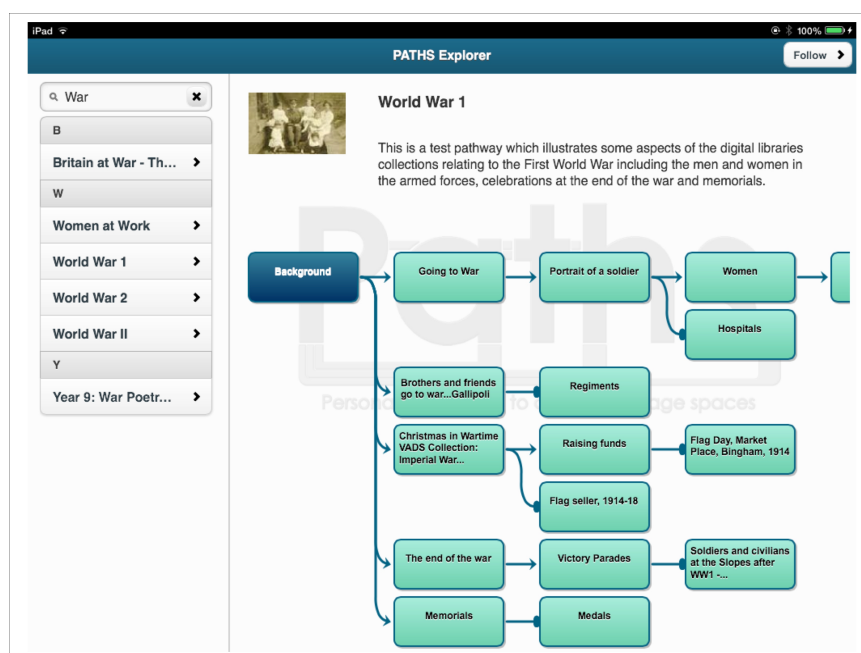
PATHS recommender system

5.7 Mobile application

A mobile application has been developed based on the iPad platform with the aim of demonstrating the usefulness of the PATHS components for users of mobile devices. A “lite” application was developed including the path following and search functionalities typically offered by cultural institutions on mobile platforms. Path creation functionality could be added in future if this is desirable.

The initial screen enables users to search for paths, to view a filtered list of the available paths and to preview a path (see below). The follow-a-path screen provides a clickable overview of the path and also allows users to view information about the content nodes.

The application is currently being prepared for deployment in Apple's App Store.



PATHS mobile application

6 Semantic Enrichment of Cultural Heritage Content

The aim of the PATHS project was to enable exploration and discovery within cultural heritage collections. To support this the project developed a range of enrichment techniques to augment the metadata for these collections with additional information to enhance users' browsing experience. This section summarises the semantic enrichment techniques developed in PATHS, with particular reference to their application to the Europeana data used in the PATHS prototype systems.

6.1 Data

The project was provided with three collections from Europeana: 1) CultureGrid, 2) Cervantes Biblioteca Virtual Miguel De Cervantes and 3) The Biblioteca Nacional de Espana. Alinari 24 ORE provided a further small collection for the project to use in its research. These collections were chosen since they contain metadata in the two languages being explored in the PATHS project (English and Spanish) in addition to containing more than 2 million items; a reasonable number for this type of research.

6.2 Motivation for Semantic Enrichment

Analysis of the three data sets revealed that adding additional information would support exploration and discovery in digital library systems and would also be helpful for interpretation of the items within them. In particular, two issues with the meta-data were identified:

- **Limited information.** Many of the items have only limited information associated with them, for example a single word title and no description.
- **Incompatible indexing schemes.** Items in Europeana are indexed using a range of hierarchical structure (e.g. Library of Congress Subject Headings and the Art and Architecture Thesaurus) that are not compatible with each other. In addition, many items are not indexed at all. Consequently there is no single index that covers all items, even for a single language.

6.3 Semantic Enrichment techniques

A range of Natural Language Processing techniques were developed for cultural heritage data and applied to the collections used in PATHS. The aim of these techniques was to provide users of cultural heritage collections with additional information that would help them to navigate and interpret the collections.

6.3.1 Identification of Key Entities

Descriptions of items often mention key entities, such as people, locations and dates. There were identified automatically in the meta-data of the three collections by applying FreeLing (Padro et al, 2010), an open source library of language processing tools which carries out several stages of linguistic analysis: identification of nouns via part of speech (PoS) tagging. Lemmatization, multiword-unit recognition and recognition of named entities (dates, places, people, organisations etc.)

6.3.2 Item Similarity

Information about which items are similar within a collection is useful for navigation, grouping together related content and recommendation of interesting content. The PATHS project developed techniques for determining the similarity between items in cultural heritage collections using Latent Dirichlet Allocation to discover latent “topics” within the collection (the approach is described by Aletras et al. (2012)).

The similarity between all items in the three Europeana collections was computed in a pairwise fashion and the 25 items with the highest score are retained for each item. This allows users to quickly navigate from one item to other related ones in the collection, even if they are from a different content provider.

6.3.3 Typed Similarity

The approaches to identifying similar items were extended to provide information about the reason pairs of items could be considered similar. This additional knowledge assists users in their understanding about how items in the collection are related together. Various types of similarity were identified: similar author, similar people involved, similar time period, similar location, similar events, similar location and similar description. Similar pairs of items were identified using a range of techniques described in Agirre et. al. (2013). The majority of these were based on comparison of the text in the relevant fields of item’s meta-data, for example the <dc:Creator> field was used to identify similar authors. Other types of similarity, e.g. similar people involved and similar location made use of the named entities that has been identified in an earlier stage of enrichment.

6.3.4 Background Links

The information associated with each item was augmented by providing links to Wikipedia. These were generated using Wikipedia Miner (Milne and Witten, 2008). See Fernando and Stevenson (2012) for further details about how Wikipedia Miner was adapted for cultural heritage documents. This process generated in-line links in the item’s meta-data. In addition, items were also mapped to Wikipedia when appropriate articles could be found, see Agirre et. al. (2012).

6.3.5 Hierarchies

The Wikipedia background links added to the item meta-data were used to automatically generate hierarchies that cover the entire collection. Two approaches are used to generate hierarchies, WikiFreq and WikiTax. WikiFreq uses Wikipedia link frequencies across the Europeana collection to organise the items. The links in the meta-data associated with each item are ordered based on their frequency in the entire collection and that set of links then inserted into the hierarchy. The WikiTax approach uses the Wikipedia Taxonomy (Ponzetto and Strube, 2011), a taxonomy derived from Wikipedia categories. Europeana artefacts are inserted into this taxonomy using the links added by Wikipedia Miner with each artefact being added to the taxonomy for all categories listed in the links. This leads to a taxonomy in which artefacts can occur in multiple locations. The two approaches are combined to create the WikiMerge hierarchy. See Fernando et al. (2012) for further details.

6.4 Application of Semantic Enrichment

The semantically enriched data from Europeana was used in the PATHS demonstration system. The similar items and background links (see 6.3.2, 6.3.3 and 6.3.4 above) were displayed to the user when they viewed an individual item. The hierarchy (see 6.3.5) was used to provide high-level navigation of the collection and was displayed in a variety of ways, including using a standard thesaurus-type view, a tag cloud and a map. See Agirre et. al. (2013) for further details about how semantic enrichment is used in the demonstrator.

6.4.1 Semantic Enrichment Using Open Source Software

Most of the enrichment techniques applied in the PATHS project were developed in-house and are relatively complex. An overall set of recommendations for the automatic enrichment of cultural heritage collections using open source software is presented in a project report available from the PATHS project web site: Agirre and Otegi (2013).

6.4.2 Representation of Semantic Enrichment using Europeana Data Model

The semantically enriched cultural heritage data in PATHS is encoded using the ESEPaths format, which is derived from Europeana Semantic Elements (ESE) and adds the enrichment information described above. ESE is the metadata scheme used to describe cultural heritage objects in Europeana (see <http://europeana.eu/schemas/ese/>). However, Europeana is moving from ESE to a new data representation, called Europeana Data Model (EDM) (Doerr et al., 2010). A project report describing how to represent the semantically enriched data can be representing using the EDM schema will be made available on the project website: (Sora et. al, 2013).

6.5 Web service for semantic enrichment

All the semantic enrichment has been done offline in PATHS. However, the project provides a web service prototype that allows independent content providers to enrich their cultural heritage items online. Specifically, the service enriches the items with two types of information: links to similar items within the PATHS collection and links to Wikipedia articles which are related to it. The Web service is described by Agirre et. al. (2013).

The web service can be accessed via http://ixa2.si.ehu.es/paths_wp2/paths_wp2.pl

7 Evaluation and field trials

7.1 Gathering User Requirements

The project began by gathering user requirements. As part of this exercise existing path-based systems used in the Cultural Heritage sector were studied (Goodale, et al, 2011). We also interviewed curators working in the culture sector about their current and potential uses of path-based content.

These studies revealed different interpretations of the path metaphor ranging from hypertext trails defined by users and exploratory navigation to processes defined by system navigation (Goodale, et al, 2012). User profiles and a user interaction model were developed, and the functionality needed in the system was identified through the user requirements analysis.

7.2 User evaluation

Both the first and second prototypes of the PATHS system have been demonstrated to and evaluated by a wide range of users in both natural and controlled task-based settings. Users were recruited to take part in trials from the UK and Ireland, Spain, Italy and Greece.

We sought participants who fulfilled one of four profiles identified during the initial user requirements analysis: experts, non-experts, teachers and students, and consumers of cultural heritage information. Use cases established for each profile helped to inform the tasks, queries and demonstration activities performed in the trials.

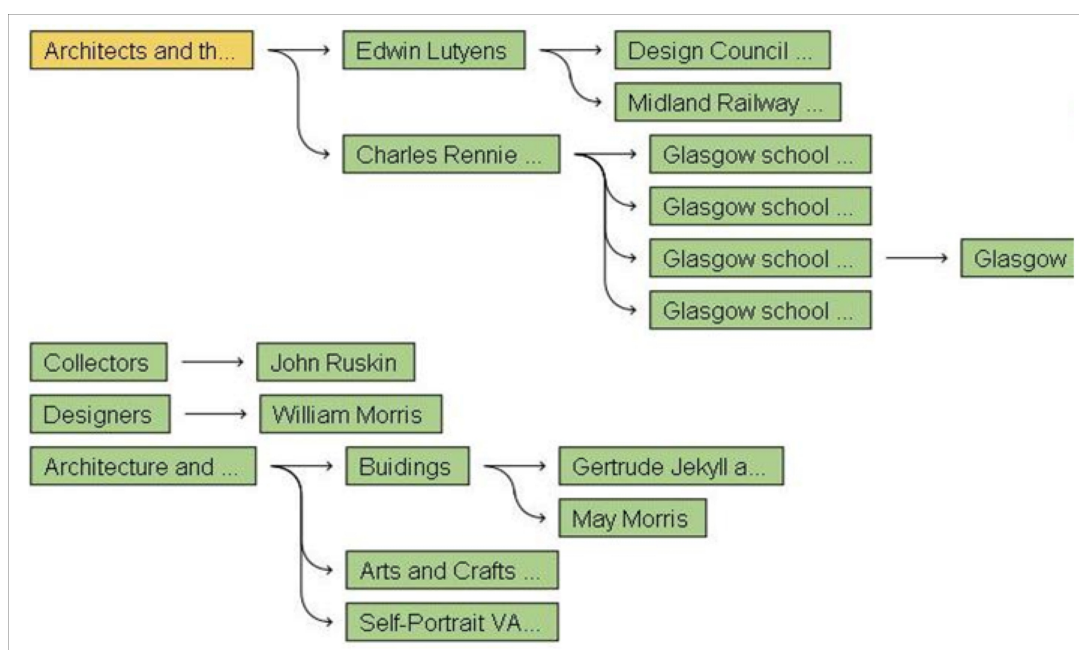
For example we asked: *“Imagine you are a student who has been asked to create a path as part of a university assignment. You have been asked to use primary source materials to create a mini online exhibition suitable for a target group within the general public and/or school visitor categories. Your goal is to introduce a historical or art-focussed topic in a popular, accessible way, and to encourage further use and exploration of cultural heritage resources.”*

The evaluation methodology used covered demonstrations, laboratory testing, field trials and also system testing of the prototypes. More than 20 demonstrations were carried out to focus groups in cultural institutions involving 108 people. 65 people took part in evaluation sessions at the iLab at the University of Sheffield completing tasks in controlled and monitored conditions. We captured in-depth feedback from participants via questionnaires and interviews on the usability and their satisfaction with the system. A final round of trials of the prototype and mobile application involved 34 people using PATHS at home for an extended period and completing diaries of their experience, which coupled with questionnaires, interviews and focus groups provided us with invaluable feedback.

Both the first and second PATHS prototype systems were well received with participants finding them mostly *Interesting, Understandable, Easy, Useful* and *Fast, Organised, Creative* and *Likeable*. The system was also seen as offering novel functionality that could be useful in a number of different user scenarios. Exploration and support for tasks were developed and extended considerably for the second prototype in light of findings and results from participants of the first prototype and now show high levels of positive responses for tasks related to *Serendipity and discovery, Finding items on a topic, Exploring*

a collection, *Sharing content with others* and *Creating resources*. The path following task, *Using content created by others* received no negative responses. It is particularly pleasing to note that responses to the second prototype improved substantially across all tasks in comparison the first prototype.

The most popular tasks were *Exploring the collection* and *Creating resources*, *Serendipity and discovery*, *Sharing content with others* and *Finding items on a topic* all ranked amongst the top three uses for PATHS.



A complex branching path created by a user during trials

The types of paths created by novice and expert users were analysed. The results suggested that subject experts are more likely to construct shorter, more focused paths than novices. With novices being more likely to construct simple linear paths, whilst experts are more likely to create complex branching paths.

PATHS is viewed as an application that is best for *searching*, *exploring* and *creating resources* by individuals using cultural heritage collections. It allows users to create personalised stories within cultural collections and, if following a path from a cultural organisation, offers meaningful guidance about the interpretation of cultural works.

7.3 Applications for Paths in Digital Cultural Heritage

We interviewed potential users about likely applications for PATHS in real world environments. Several core use case scenarios emerged:

- **Cultural heritage:** collection overviews, online exhibitions, guided tours, creative works derived from CH objects
- **Teaching and learning (teacher):** lesson plans and learning objects, demonstrating good practice in information tasks
- **Teaching and learning (student):** group work, inquiry-based learning activities
- **Research:** keeping a history of items found in a collection, keeping notes, sense-making, discussing with colleagues
- **Community:** co-creation, sharing and discussing items and paths
- **Leisure:** local history research, keeping track of favourite items, discussing with other enthusiasts, sharing with friends and family

Applications outside the cultural heritage domain have also been identified, including professional information work in areas such as law and health.

8 Summary of activities

A summary of the activities that were carried out in PATHS is as follows:

- **State of the art monitoring** – the project has been monitoring the state of the art in Educational Informatics, Information Retrieval and Semantic Similarity relatedness.
 - Long term research in the Information Retrieval community that is attempting to see search queries and the presentation of results from different user perspectives has been of particular interest to PATHS.
 - Partners took part in the Cultural Heritage in Context exercise used Europeana data to evaluate both information retrieval and semantic enrichment generating concepts from Wikipedia and the Linked Open Data cloud.
 - Advances in methods to calculate the Semantic Textual Similarity between sentences in documents and in Named Entity Recognition and Classification are also contributing to improved performance.
- **Sentiment analysis** – the project carried out analyses of sentiment towards Europeana and to cultural heritage content online. The sentiment towards Europeana expressed online between its launch on November 2008 and February 2011 showed the negative impact of system issues that marred the launch of the service. Renewed analysis of sentiments from June 2011 to June 2012 shows that the negativity has now been forgotten and that more of the discussion about Europeana is being held in mainstream media rather than in specialist blogs. Special exhibitions, such as the one focused on the 1914-18 war, were a particular driver for neutral and positive stories about Europeana.
- **User evaluation** – the project has completed a comprehensive programme of user evaluation that began with the first prototype, continued with evaluation of the second prototype and concluded with field trials of the final PATHS prototype and mobile application. We aimed to engage with participants who represented the target end-user communities of the PATHS system and involved them in a series of demonstrations, laboratory trials and field trials to gather their feedback and to inform the research and development cycle.
- **Establishing prototypes** – the project developed a first prototype of the PATHS system based on the user requirements analysis and functional specification. This prototype was evaluated by users and subjected to system testing, the results informing the development of a second PATHS prototype. A mobile application for use on iPads has been developed along with additional applications including a recommender system and a prototype API for enrichment of content items on the fly.
- **Content analysis and enrichment** – the project has analysed and enriched more than 2 million items of content from three Europeana collections (CultureGrid, Cervantes Biblioteca Virtual Miguel De Cervantes and the Biblioteca Nacional de Espana) and from Alinari 24 Ore. The content was formatted and augmented with additional information to improve its utility in search interfaces by adding index terms, creating weighted links between items, adding background links and adding vocabulary terms linked to a subject hierarchy.

- **Personalisation and recommendations** – the potential to personalise users' experiences in cultural heritage information systems has been explored through analysis of query logs, study of user profiles and the similarity of content items. A recommender service was implemented in the second PATHS prototype.
- **Extensions to ESE and EDM** – the metadata used in PATHS was provided in ESE format (Europeana Semantic Elements). The project developed extensions to ESE to capture the enrichments produced by the Natural Language processing carried out in the project. The project team contributed to the development of EDM (the Europeana Data Model deployed by Europeana in 2012) by transforming its ESE-PATHS format to EDM-PATHS.
- **Dissemination and awareness raising** were important activities in PATHS.
 - Partners were active in publishing papers and presenting results to the both the research community and to the cultural heritage sector.
 - Partners were active members of the technical networks and the Europeana professional network. Several partners participated in Europeana hackathons and in SemEval challenges.
 - The project website www.paths-project.eu was established and has been continuously developed throughout the project as reports, publications and presentations have been published.
 - Project news and activities have been disseminated to the wider network through a regular newsletter and other channels via Twitter, LinkedIn and mailing lists.
 - The organisation of workshops under the framework of TPDL during years two and three have been key activities for PATHS providing opportunities to bring together researchers and cultural heritage institutions to discuss ways of supporting user experiences of digital libraries.
 - In addition to these activities PATHS partners have presented the project's results at a number of national and international conferences to both the research community and to the cultural sector.

9 Impact and exploitation

PATHS set out to research and develop techniques to improve access to very large digital library collections being offered to end-users.

The following summarises the project's main areas of impact:

- Demonstrating the potential to enrich the content being provided by cultural heritage to improve discovery of quality content items by users.
- The PATHS prototype has demonstrated the potential to improve the quality of users' experiences of using very large digital libraries by offering novel visualisations of the content space and providing interactive tools to create and use paths.
- PATHS has been a pioneering project within the Europeana network contributing to the development of the Europeana Data Model and demonstrating new techniques for enriching content.
- PATHS engaged with 207 participants from over 17 organisations related to cultural heritage and education, across six different European countries in its evaluation activities.
- More than 30 research publications have been produced by project partners.
- The PATHS market study contributes to understanding of the market for software and services for information management and retrieval amongst cultural institutions.
- Developing a versatile set of components and methodologies that can be adopted and exploited inside existing products and services. Projects such as LoCloud are leveraging PATHS results by implementing them in micro-services delivered through the cloud.
- A prototype web service that enables cultural institutions to enrich their cultural heritage items with links to related items and to Wikipedia pages has been taken up by the LoCloud project for development.
- PATHS information retrieval components have been exploited by AVINET in their web GIS portal Adaptive.



Participants at the SUEDL workshop, 2012

Exploitation

Through its work, PATHS has supported the consortium members in their future exploitation of PATHS results. The project has prepared an exploitation plan based on a market survey of the business needs of organisations that are potential users of PATHS technology and the partners' visions and ideas.

Three alternative strategies for sustainability and exploitation have been identified:

Scientific exploitation of knowledge and prototype software products that will be sustained through continued research; this includes projects that need further work before they are ready for the market. The following 'knowledge products' have significant potential for scientific exploitation:

- The PATH paradigm to be refined, standardized and adapted to handle any content through a Linked Data approach;
- Information retrieval services;
- Natural language processing techniques;
- Map based information visualization and exploration of thesauri.

Commercial exploitation of PATHS products that constitute commercial software and services and PATHS function libraries or APIs with potential to constitute part of existing or new commercial software products. The following products have significant potential for commercial exploitation:

- Recommender systems;
- Open Web API for application development;
- Content enrichment through Natural Language Processing techniques as a Cloud SaaS service;
- Sentiment analysis tools for web content.

Open source development of PATHS products as either stand-alone components or as new modules of existing, widely deployed Open Source software products with a significant developer base and interest community.

More information about the PATHS products and their availability can be found on the project website at the address below.

10 Further information

A video promoting PATHS results is freely available at: <http://www.paths-project.eu>

11 References

- Agirre, E., A. Barrena, O. Lopez de Lacalla, A. Soroa, M. Stevenson and S. Fernando. 2012. Matching Cultural Heritage items to Wikipedia. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 1729--1735, Istanbul, Turkey
- Agirre, E., P.D. Clough, S. Fernando, M. Hall, A. Otegi and M. Stevenson. 2012. *The Sheffield and Basque Country Universities Entry to CHiC: using Random Walks and Similarity to access Cultural Heritage*. CLEF 2012.
- Agirre, E., N. Aletras, P. Clough, S. Fernando, P. Goodale, M. Hall, A. Soroa and M. Stevenson. 2013. PATHS: A System for Accessing Cultural Heritage Collections. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 151--156, Sofia, Bulgaria
- Agirre, E., N. Aletras, A. Gonzalez-Agirre, G. Rigau and M. Stevenson. 2013. *UBC UOS-TYPED: Regression for typed-similarity*. In *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, Volume 1: *Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, pages 132--137, Atlanta, Georgia, USA
- Agirre, E., A. Barrena, K. Fernandez, E. Miranda, A. Otegi and A. Soroa. 2013. *PATHSenrich: A Web Service Prototype for Automatic Cultural Heritage Item Enrichment*. In *TPDL 2013, Lecture Notes in Computer Science volume 8092*, pages 462--465.
- Agirre, E., D. Cer, M. Diab, A. Gonzalez-Agirre, W. Guo. 2013. **SEM 2013 shared task: Semantic Textual Similarity*. *SEM 2013.
- Agirre, E. and A. Otegi. 2013. *Recommendations for the automatic enrichment of DL content using open source software*. PATHS Project Report. Available via <http://www.paths-project.eu/eng/Resources>
- Aletras, N., M. Stevenson, and P. Clough. 2012. Computing similarity between items in a digital library of cultural heritage. *Journal of Computing and Cultural Heritage*, 5(4):no. 16.
- Clough, P.D., A. Otegi, E. Agirre and M. Hall. Sep 2013. Implementing Recommendations in the PATHS System. SUEDL 2013.
- Fernando S. and M. Stevenson. 2012. Adapting Wikification to Cultural Heritage. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 101--106, Avignon, France.
- Fernando, S., M. Hall, E. Agirre, A. Soroa, P. Clough, and M. Stevenson. 2012. Comparing taxonomies for organising collections of documents. In *Proc. of COLING 2012*, pages 879--894, Mumbai, India.
- Fernando, S., P. Goodale, P.D. Clough, M. Stevenson, M. Hall and E. Agirre. 2013. *Generating Paths through Cultural Heritage Collections*. LATECH 2013.
- Fernie, K., J. Griffiths, P. Archer, K. Chandrinou, A. de Polo, M. Stevenson, P. Clough, P. Goodale, M. Hall, E. Agirre, O. Lopez de Lacalle, R. Bergheim. 2012. PATHS: Personalising access to cultural heritage spaces. In: *Proceedings of the 18th International Conference on Virtual Systems and Multimedia (VSMM 2012)*, 469-474. IEEE: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6365960&tag=1
- Hall, M., O. Lopez de Lacalle, A. Soroa, P. D. Clough and E. Agirre. 2012. *Enabling the Discovery of Digital Cultural Heritage Objects through Wikipedia*, LaTeCH 2012.
- Hall, M., P. Clough, S. Fernando, P. Goodale, M. Stevenson, E. Agirre, A. Otegi, A. Soroa, K. Fernie and J. Griffiths. 2013. *Information Seeking in Digital Cultural Heritage with PATHS*. SIGIR 2013.

P. Goodale et al. (2012). User-Centred Design to Support Exploration and Path Creation in Cultural Heritage Collections. In: *Proceedings of EuroHCIR2012*, (Nijmegen, Netherlands, August 25, 2012), 75-78.

Intxaurren, A., O. Lopez de Lacalle and E. Agirre. 2011. *UBC at Slot Filling TAC-KBP 2011*, Nov 2011.

Soraa, A., E. Agirre and A. Otegi (forthcoming) RoadMap from EEPATHS to EDMPATHS: a user case. PATHS Project Report. <http://www.paths-project.eu/eng/Resources>

Yeh, E. and E. Agirre. 2012. *SRIUBC: Simple Similarity Features for Semantic Textual Similarity*. *SEM 2012.

11.1 Project Deliverables

Public project deliverables are available from the PATHS project web site: <http://www.paths-project.eu/eng/Resources>.

D1.1 User Requirements Analysis by Paula Goodale, Mark Hall, Kate Fernie and Phil Archer.

D1.3 Functional Specification of the first PATHS prototype by Phil Archer with Mark M. Hall, Paul Clough, Mark Stevenson, Eneko Agirre, Iñaki Alegria and Kate Fernie.

D1.4 Final State of Art Monitoring Report by Phil Archer, Konstantinos Chandrinou, Mark Stevenson, Mark M. Hall, Paul Clough, Paula Goodale, Eneko Agirre, German Rigau, Nigel Ford, Jillian Griffiths with Kate Fernie, Timos Kouloumpis.

D1.5 Functional Specification for second prototype by George Chrysochoidis, Phil Archer, Kostas Chandrinou, with Mark Hall, Paul Clough, Paula Goodale, Mark Stevenson, Eneko Agirre, Aitor Soraa, Iñaki Alegria, Jillian Griffiths, Kate Fernie.

D2.1 Processing and representation of Content for the first prototype by Eneko Agirre and Oier Lopez de Lacalle with Aitor Soraa, Mark Stevenson, Samuel Fernando, Nikos Aletras, Antonis Kukurikos, Kate Fernie.

D2.2 Processing and representation of Content for the second prototype by Eneko Agirre, Arantxa Otegi and Aitor Soraa with, Nikos Aletras, Constantinos Chandrinou, Samuel Fernando, Aitor Gonzalez-Agirre

D3.1 System architecture specification by Stein Runar Bergheim and Idar Thoresen Kvam with Phil Archer, Kate Fernie, Paul Clough, Tor Gunnar Øverli and Mark Stevenson.

D3.2 First PATHS Prototype by Stein Runar Bergheim, Mark Hall, Eneko Agirre, Aitor Soraa, Antonis Kukurikos, Kate Fernie, Tor Gunnar Øverli.

D3.3 Second Prototype and Documentation by Stein Runar Bergheim, Mark Hall, Eneko Agirre, Kate Fernie and Alok Singh.

D4.2 Final Prototype Interface Design by Mark Hall, Paula Goodale, with Paul Clough, Eneko Agirre, Kate Fernie, Jillian Griffiths, Mark Stevenson.

D4.3 Additional Applications by George Chrysochoidis, Paul Clough, Eneko Agirre, Arantxa Otegi and Kate Fernie.

D5.1 Evaluation of the first PATHS Prototype by Jillian Griffiths, Paula Goodale, with Sam Minelli, Andrea de Pollo, Rodrigo Agerri, Aitor Soroa, Mark Hall, Stein Runar Bergheim, Konstantinos Chandrinou, George Chryssochoidis, Kate Fernie, Tom Usher.

D5.2 Evaluation of Second PATHS Prototype by Jillian Griffiths, Sheena Bassett, Paula Goodale, Rodrigo Agerri, George Chryssochoidis, Kate Fernie and Jen Smith.

D5.3 Field Trials of the PATHS Prototype by Jillian Griffiths, Sheena Bassett, Paula Goodale, Jen Smith, Eneko Agirre, Arantxa Otegi, Stein Runar Bergheim, George Chryssochoidis, Kate Fernie and Samuel Fernando.

D6.3 Exploitation Plan by Frode Skaevestad, Runar Bergheim, Kate Fernie, Mark Stevenson, Eneko Agirre, Arantxa Otegi, Jillian Griffiths, Robert Davies, Tor Gunnar Overli, George Chryssochoidis and Konstantinos Chandrios.