Grant Agreement No. **ICT-2009-270082**

Project Acronym **PATHS**

Project full title **Personalised Access To cultural Heritage Spaces**

D 6.4

Exploitation Plan

Authors: Frode Skjævestad (Avinet)
Runar Bergheim (Avinet)

Contributors: Kate Fernie (MDR Partners)
Mark Stevenson (USFD)
Eneko Agirre (UPV/EHU)
Arantxa Otegi (EPV/EHU)
Jillian Griffiths (MDR Partners)
Robert Stanly Davies (MDR Partners)
Tor Gunnar Øverli (Avinet)
George Chryssochoidis (iSieve)
Konstandinos Chandrinos (iSieve)

Project funded under FP7-ICT-2009-6 Challenge 4 – "Digital Libraries and Content"

Status	Final
Distribution level	Public
Date of delivery	17/01/2014
Type	Report
Project website	http://www.paths-project.eu
Project Coordinator	Dr. Mark Stevenson University of Sheffield

Change Log

Version	Date	Amended by	Changes
0.1	02-01-2014	Frode Skjævestad	First draft
0.2	10-01-2014	Frode Skjævestad	Updated based on input from partners
0.3	15-01-2014	Frode Skjævestad	Updated based on input from partners
1.0	16-01-2014	Frode Skjævestad	Updated chapter 4.3.3

Table of Contents

List of figures

1 Executive summary

The aim of this report is to elaborate the market potential and how to exploit the knowledge, research results, software, services and methodologies of the PATHS project. In this report, 'products' is used as a collective term to describe all exploitable outcomes of the PATHS project. A market study has been carried out within the Europeana community and the results are presented in the report.

All PATHS products are defined and described by the responsible organization. In order to point out possible ways of making the products marketable after the end of the project period the description of each product is important.

IPR constraints on both project level and partner organisation levels are elaborated.

Commercial rights related to each of the products have been clarified and the cost/profit model established

An online questionnaire was executed towards a sample of organizations that are envisaged to be relevant end-users of PATHS products and services. The results of the survey have been analysed in order to determine how best to sustain each of the products.

Depending on the nature of the 'product', a strategy has been formulated to secure continuation and uptake in the market. This is achieved through three alternative strategies for sustainability and exploitation:

Scientific exploitation is applicable to knowledge and prototypical software products that will be sustained through continued research. It also applies to projects and that need further work before they are ready for the market. The following 'knowledge products' have been found to have significant potential for scientific exploitation:

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

The second exploitation strategy is **commercial exploitation**, i.e. the PATHS products that independently constitute commercial software and services, or the PATHS function libraries that constitute part of other existing or new commercial software products. The following products have been found to have significant potential for commercial exploitation, i.e. to constitute independent or partial commercial software and/or services:

- 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.

PATHS products that do not fit into either of the first two strategies but have significant re-use value will be sustained through a strategy that will transform the products into stand-alone **Open Source** components or commit new modules to existing, widely deployed Open Source software products with a significant developer base and interest community.

Action plans have been prepared that defines specific follow-up activities for the prioritized products. In addition to describing the necessary steps forward, the plans also assign responsibilities to legal persons to secure the timely execution of the exploitation plan after the completion of the project.

2 Introduction

2.1 Overview of Exploitation in PATHS

Exploitation has been on the agenda of PATHS from the first months of the project, related to the development of tool, services and prototypes, user evaluations, dissemination activities and marked assessments.

Two functional prototypes have been developed and evaluated, altogether this work forms a large pool of competence, knowhow, experience and operational tools and components. For exploitation, re-use of these results in scientific, commercial and educational purposes is required.

PATHS has been contributing to the Europeana initiative with technology, data integration and promotion, and enabling low-cost integration through APIs for cultural portals. The results are being made available as open-source components whenever this is applicable.

Exploitation of PATHS project results in the post-project phase is described and planned in this document. The document is based on a marked survey of business needs from organisations that are potential users of the PATHS technology and from the partners' own visions and ideas for re-use of the results of PATHS.

Protection of intellectual rights is critical for correct exploitation and re-use and even though the principles of open-source is being given priority, IPR is a major element of the exploitation plan and is resulting in a set of license and distribution agreements.

2.2 Roles and Responsibilities

The development of the exploitation plan has been a joint effort of the entire PATHS consortium. AVINET has been responsible for developing the structure of the exploitation plan, for conducting the market study as well as the review and analysis of inputs contributed by the wider consortium.

MDR, in the company's capacity as project manager, has been responsible for clarifying and developing the license and distribution agreements that makes it possible to move forward with the joint and separate exploitation strategies for each of the PATHS products.

Finally, all partners contributed to the development of the action plans and have been assigned joint or independent responsibilities for implementing the exploitation plan depending on the nature of the specific product and activity.

2.3 Strategy and Action Plan

In order to keep the PATHS knowledge, products and services alive after the end of the project it is necessary to have a market strategy and an action plan describing how to put the project results further. A strategy in this sense will be a plan or method for achieving our goal over time. An action plan describes how we want to go further with the PATHS products. Every product group is listed in chapter 9 together with a vision for

how they could be brought to market, an named individual with responsibility for exploring this and an expected timescale.

3 Methodology

This chapter describes the methodology that has been applied in the preparation of the exploitation plan and its underlying components.

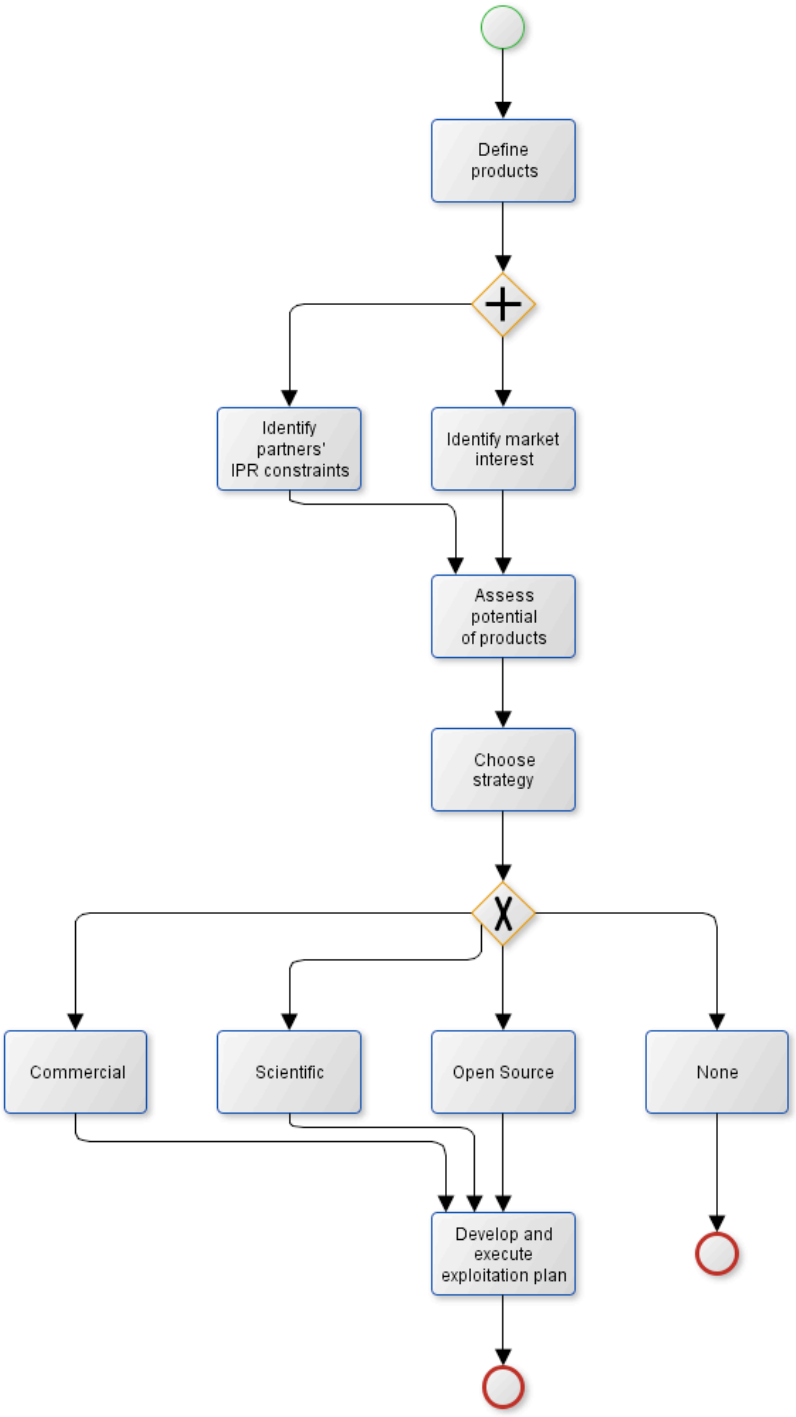


Figure 1: Schematic presentation of Methodology

3.1 Define products

The scientific research in the PATHS project produces three distinctive types of results that in this document are referred to as “products”. The term is used to denote “something” that can be transformed into sustainable, marketable entities:

1. Research results & knowledge
2. Software
3. Services & methodologies

In order to describe how these products may be brought to market through the exploitation activities of the PATHS project it is first necessary to identify the distinctive project results that satisfy the above criteria, bundle them and classify them according to the above categorization.

This is dealt with in Chapter 4.

3.2 Identify partners’ IPR constraints

Having successfully identified all the “products” resulting from the project it is necessary to identify the possibilities and constraints posed by the IPR policies of:

A. The consortium agreement, project level agreements
B. The individual partners IPR policies
C. Base licenses with terms and conditions that propagate upstream

This is dealt with in Chapter 5.

3.3 Identify market interest

Once the products are identified and IPR issues have been cleared, it is necessary to identify the market interest and permissible business models for PATHS products within the envisaged target market.

For this purpose, a market survey was conducted to study the business needs of organisations who are potential users of PATHS technology, any barriers to adoption, competitive products and services as well as the potential market size and development over a 5 year period.

3.4 Assess potential of products (SWOT)

Review the potential for the development of commercial products or services based on PATHS’ results, alternative exploitation strategies, risks and threats, and the potential routes to market.

3.5 Choose strategy and develop exploitation plan

An exploitation plan has been developed for the project based on the market survey and the review of development potential. The exploitation plan details the exploitable results of the project, how they can impact the research field and enter the market strategies. Its aim is to explain how the outputs of the project will be exploited and the common and individual strategies of project partners. The document will be maintained and updated to reflect possible refocusing of user-oriented and industrial work.

3.6 Interaction

All partners have been involved in preparing a Market Agreement for the Consortium stating the IPR and rights for using technology and content within the consortium during the project's lifetime. MDR will develop license and distribution agreements for the project results. This is dealt with in Section 9.

4 PATHS Products

This chapter identifies PATHS products and services. The partner or partners leading the development of each product is listed in parenthesis following its description.

4.1 Research results, knowledge

4.1.1 PATHS as a paradigm for navigating cultural heritage collections (all)

The PATHS project made use of paths, or trails, through cultural heritage collections as a navigation paradigm. The use of trails to navigate collections is not a new approach, and was described by Vannevar Bush as far back as 1945 in the context of the MEMEX system. However, previous applications have focussed on the development of tools that support path creation on open, dynamic collections such as the internet (e.g. Walden's Paths and Storify) or on static trails through collections that are effectively tours and these are normally manually created (e.g. Europeana's virtual exhibitions). In addition, little investigation has been undertaken regarding the kinds of paths users create to guide users through digital collections on specific topics or themes.

PATHS offers users functionality in which they can follow, create and share paths through a 'closed' collection of cultural heritage items. An advantage of the choice of using a closed collection is that a significant amount of processing can be applied to enrich the collection and this supports the user's exploration and path creation. The same processing could not have been applied to an 'open' collection, such as the Internet, due to its size, heterogeneity, and constant change. A further advantage of the approach used in the PATHS project is that by allowing users to create their own paths through the collection, it allows the user to explore the collection more fully and dynamically than would be possible if they were following fixed trails.

4.1.2 Information Retrieval (USFD, UPV/EHU)

The project has contributed to existing knowledge about functionality that can support users with exploratory searching and sensemaking activities. There is currently much interest in supporting exploration (as compared with more traditional searching and browsing) and the novel visualisations, hierarchical navigation functionality, paths and workspace feature are all ways of supporting users with explore activities. The project has investigated the role of users' cognitive style on their navigation patterns, another area currently under investigation in IR research.

4.1.3 Recommender Systems (USFD, UPV/EHU)

The project experimented with various forms of recommendation to support users during exploration and navigation of the digital content. Recommendations utilising content-based and concept-based approaches have been experimented with during the project to provide personalised and non-personalised (or generic) recommendations for specific items. The use of typed similarity methods is one example of how state-of-the-art research has been advanced in the PATHS project. Additional experiments have also been carried out making use of different sources of information to provide recommendations. This is on-going research and all experiments and results will be published in a forthcoming research publication.

4.1.4 Natural Language Processing (UPV/EHU, USFD)

The project applied a range of Natural Language Processing (NLP) techniques to cultural heritage content in order to enrich it with additional information that supports the user's experience of navigating the collection. This process included the direct application of existing NLP tools without modification, the tuning of tools to the cultural heritage domain and the exploration of novel problems. The choice of approach depended on whether appropriate software tools and techniques were available or not. The consortium sought to re-use existing tools where these were available, although novel algorithms and techniques were also developed during the project. More specifically, the following NLP techniques for the analysis of cultural heritage content were developed:

1. Identification of key entities within item descriptions. This was carried out by applying existing NLP toolkits (Freeling and GATE) to collection metadata. Performance of each toolkit was compared and Freeling was used due to its reliable analysis and support for multiple languages.

2. Generation of internal links. Algorithms were developed to identify the similarity between items in the collection based on their meta-data. These approaches were applied on a large-scale (to around half a million items) in order to identify pairs of similar items. A novel extension of this work was the development of techniques to identify the type of the similarity, this is in contrast to the majority of work on similarity that has been carried out within NLP which has been concerned with generic notions of "similarity" and "relatedness".

3. Generation of typed internal links. PATHS also identified additional types of similarity relevant to the cultural heritage collections used in the project (e.g. "similar author" and "similar time period"). The identification of these similarity types was motivated by users who wanted to be provided with a reason for why items were similar in order to enrich their navigation experience and understand the reason why items are being recommended to them. The typed similarity task was used as part of a community evaluation exercise at a major international conference.

4. Generation of links to Wikipedia articles. Links to related material were added to the collection metadata in order to provide background information about the items and additional explanations about their significance to users. These were inline links connected with important terms mentioned in the item's description. They were added using the freely available WikiMiner software tool developed by the University of Waikato. WikiMiner's output was post processed to generate as good links as possible for cultural heritage material.

5. Generation of identity links to wikipedia articles. In addition, further links were generated by linking entire items to relevant Wikipedia entries where available. These links were not inline, in the sense that they do not annotate the item's metadata, and are simply attached to the item.

6. Automatic generation of hierarchies. Hierarchies are useful to support navigation and provide the user access points to and an overview of the collection. However, the aggregated collections from Europeana used in the project do not use a single hierarchy, in fact some of the collections do not provide any hierarchical information. Consequently, algorithms to automatically create hierarchies from the collection were developed. The first algorithm, WikiFreq, uses Wikipedia link frequencies across the Europeana collection to organise the items. The second, WikiTax, uses the Wikipedia Taxonomy, a taxonomy

derived from Wikipedia categories. The two approaches are combined to create the WikiMerge hierarchy.

4.1.5 Information Visualization (USFD)

Within the PATHS project a novel 'map-based' interface has been developed that provides an overview of the semantic space. This advances existing understanding of methods for providing collection overviews and the map-based interface in PATHS combines methods from Geographic Information Systems (GIS) and Natural Language Processing (NLP). The map-based interface has formed the basis of peer-reviewed publications and will be an aspect of the project that partners will take forward to produce a more refined method for visualising cultural heritage document collections. The potentially exploitable software of the map-based interface is described in greater detail below under "software products" below.

4.2 Software products

4.2.1 Natural language processing tools for enrichment of metadata (UPV/EHU, USFD)

1. Identification of key entities within item descriptions: 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.). The software is owned by UPV/EHU. It uses Freeling (GPL v2.0 with dual license - commercial licensing also available) and Stanford CoreNLP (GPL v2.0 with dual license - commercial licensing also available).

2. Generation of internal links: Information about which items are similar within a collection is useful for navigation, grouping together related content and recommendation of interesting content. Links between items in the collection were generated based on the similarity between the items. The technique used for determining the similarity between items in cultural heritage collections was Latent Dirichlet Allocation (LDA) to discover latent "topics" within the collection. The approach is described by Aletras et al. (2012). Ownership of this software belongs to USFD. It uses the gensim package (GNU LGPL) to carry out LDA.

3. Generation of typed links: The previous software was extended to indicate 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) obtaining the second position in a public competition. Ownership of this software belongs to UPV/EHU and USFD. It uses Weka for the machine-learning component, which is GPL 3.0. Although the prototype used internal links produced above, it could use other software, including in-house software by UPV/EHU.

4. Generation of links to Wikipedia articles: The information associated with each item, which is sometimes very limited, was augmented by providing links to related articles in Wikipedia. The software used is based on 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. Ownership of this software belongs to USFD, and includes Wikipedia Miner (GPL License).

5. Generation of identity links to Wikipedia articles: In addition, software to map items to Wikipedia articles when appropriate articles could be found was produced, using the links generated in the previous step (Agirre et. al. 2012). Ownership of this software belongs to UPV/EHU and USFD. Although the prototype used links to Wikipedia produced with Wikipedia Miner, those links could be produced with other software, including in-house software by UPV/EHU.

6. Automatic generation of hierarchies: The Wikipedia background links added to the item meta-data were used to automatically generate hierarchies that the 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 metadata 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. Ownership of this software belongs to UPV/EHU and USFD. The software used a resource called WikiTax (GNU Free Documentation License). Although the prototype used links to wikipedia produced with Wikipedia Miner, those links could be produced with other software, including in-house software by UPV/EHU.

4.2.2 "Topic map", map visualization of thesaurus

The name of this product is merely a working title and should not be confused with the semantic markup technology XML Topic Maps (XTM) 1.0. In PATHS, "topic map" refers to a visualization whereby hierarchically structured semantic concepts that are attached to the documents are used to create a visualisation of the semantic space that resembles a Google Map. The PATHS approach is novel in that the hierarchical structure is exploited to enable the approach to scale to large document collections and to create a map where the higher levels of spatial abstraction have semantic meaning. For further information about the approach refer to (Hall & Clough, 2003).

The novel Hierarchical Spatialisation Algorithm (HSA) created in PATHS requires as its input a thesaurus, a collection of documents, and a mapping of the documents into the thesaurus. From these it generates a hierarchical, semantic map using a six-step pipeline. The first three steps (Tree pruning, Item pruning, Vectorisation) pre-process the thesaurus into the structure required for the main spatialisation algorithm. The next two steps (Spatialisation, Positioning) create the spatialisation of the thesaurus and the documents in the collection. The last step (Post-processing) creates the final map outline.

The output of the approach is a map-based overview of concepts used to describe the document collection that resembles a Google Map. Topics in the thesaurus are shown as islands; the size and centrality of each island is determined by the number of sub-topics, as

well as the number of information items attached to each topic node. As the user zooms further in on the map, leaf topics from the next hierarchical level of the branch become visible. Once the user reaches detailed zoom levels, information items that are attached to the leaf topics are displayed as clickable map pins. When the user clicks on a map pin, they are hyperlinked to the landing page of the individual resources.

The PATHS project developed the technology necessary to create a virtual landscape based on a multi-hierarchy of labelled topics. For the purpose of rendering this into map tiles, usable by the commonplace Internet mapping frameworks, the Mapnik library is utilized. Mapnik is Open Source under the LGPL license.

The visualization mechanism utilizes the JavaScript mapping library Leaflet to create a responsive user interface on the client side. Leaflet is also Open Source under a special license that permits modifications and redistribution provided the license is published alongside the library. In the PATHS prototype, PostgreSQL and PostGIS (both Open Source) are used to serve the topics that are shown as pins in the map. This is however not a strong dependency and it is possible to alter the product so that it can be used with other back-end technologies.

The “topic map” technology could be applied to any hierarchy in any domain, not necessarily a topic hierarchy and definitively with a wider application than the cultural heritage domain. This product is too narrow to live on its own but could be refined into a plugin to a mainstream software (i.e. Wordpress, Joomla or similar) or be made part of an existing, widely used visualization library (i.e. D3.js or similar)

4.2.3 Recommender system: real-time generation of links to relevant content in same collection (UPV/EHU, USFD)

The developed recommender system provides non-personalised recommendations based on item-to-item co-occurrences. Co-occurrence information (items that have been viewed consecutively in the same session) has been mined from a sample of Europeana portal logs to power the recommendations (Clough et al., 2013). The system assists users with recommendations of the kind "people who viewed this item also viewed this item", providing links to other items in the collection. The owners of this product are UPV/EHU and USFD.

4.2.4 Sentiment analysis for web content resources

This proprietary technology was part of existing IPR brought into the project by consortium partner i-sieve and was already at the outset a well-established commercial service. The sentiment analysis is carried out by a piece of software that takes two inputs:

1. a topic, typically a brand or the name of a physical or legal person and keywords associate with this brand
2. existing web resources, automatically identified through major search providers (i.e. Google, Yahoo) indexes

The software first identifies the relevance of each web resource against the topic that is being analyzed and then extract information on what “*the web*” says about the topic, why “*the web*” is saying what it is saying and how “*the web*” perceives a topic in relation to its counterparts (i.e. competing brands). In the PATHS project, this technology was applied to (1) the brands “Europeana” and “PATHS” versus other comparable entities and (2) individual cultural heritage content resources.

The software already has a wide customer base and through the project new application areas within the cultural heritage domain has been identified. As stated in the project agreement, this is a proprietary, non-open source technology that was brought into the project in a near complete state and is the exclusive property of i-sieve technologies.

4.2.5 Open Web API for working with PATHS

The Web API for working with PATHS is in its present form has a strong dependency towards the underlying Europeana based data model. This limits the number of application areas where it may be used from the outset. However, as both the data and software architecture was designed with flexibility in mind, functionality related to authoring, retrieval and discovery of PATHS can easily be made independent of the rest of the API, permitting easy implementation of the API on top of any mainstream database such as MySQL, PostgreSQL, Oracle and SQL Server.

The API is developed as a set of XML/JSON web services implemented using programming language C#.NET. For communication with the Solr search engine, the Solr.NET library is used, possibly to be replaced by Lucene.NET for a more compact application. The entire API is developed within the PATHS project and is in line with the project agreement Open Source and will be published on GitHub. There are no obstacles to commercial or free exploitation of this product.

4.3 Services, methodologies

4.3.1 Content enrichment (USFD, UPV/EHU, AVINET)

The content enrichment techniques developed in the project rely on a range of pre-existing tools to carry out linguistic analysis. The majority of these are freely available and could be reused but this would be difficult without significant technical expertise. For example, the similar item links make use of an implementation of Latent Dirichlet Allocation which is freely available on the internet under open source license. See description of individual products above for details on ownership and licensing.

A **content enrichment web service** prototype has been developed which allows independent content providers to enrich cultural heritage items in their own collections. The technical expertise required to use this service is significantly lower than making use of the content enrichment techniques directly. The service enriches the items with two types of information: links to similar items within the PATHS collection and links to related Wikipedia articles. The web service is described by Agirre et al. (2013b). It is owned by UPV/EHU. It uses Solr (Apache license).

4.3.2 Development of user requirements specifications

The PATHS prototypes were developed based on comprehensive user requirements specification that was developed based on a refinement of well-known methodologies. The process starts with a user requirements analysis including: desk research, surveys/questionnaires and user experiments. Based on this analysis, a functional user requirements specification was developed. The methodology was developed by USFD and is described in detail in deliverable D1.1. The methodology may be used as a model for offering (commercial) user requirements specification consultancy services subject to market response.

4.3.3 Web application user evaluation

Evaluation and Field Trials within PATHS comprised a programme that began with the evaluation of the first prototype, D5.1 Evaluation of the first PATHS Prototype (2012), continued with evaluation of the second prototype, D5.2 Evaluation of the second PATHS Prototype (2013) and concludes with field trials and technical testing reported in D5.3 Field Trials of the PATHS Prototype (2014).

A robust holistic programme of evaluation encompassing demonstration activities, laboratory testing, field trials and project-wide evaluation of technical components was developed by MDR and USFD to assess the PATHS prototypes. This included:

- a series of **demonstration focus group sessions** organised within cultural institution settings to enable quantitative and qualitative data to be collected from focus groups to assess the usability and usefulness of PATHS and its use as a tool to explore cultural heritage collections. In total 22 demonstrations were undertaken with 108 participants.
- **laboratory based evaluation** which allowed for testing of the efficiency and effectiveness of the Paths software under controlled conditions (including use of tasks and Morae software), along with in-depth feedback from test participants on usability and satisfaction. A total of 65 participants took part in these activities.
- **field trials** of the prototype and mobile application were conducted which included use a diary study approach, coupled with questionnaires, interviews and focus groups. 34 participants took part in these activities.
- **project-wide evaluation activities** informed, complemented and extended the overall user-centred evaluations.

During evaluation of both Prototypes groups of target users were invited to take part in field-based demonstration evaluation sessions, which were locally hosted sessions at cultural institutions and other venues. These sessions provided an opportunity for the collection of qualitative data from focus group discussions and quantitative feedback through questionnaires. These sessions demonstrated activities based on the task employed within the laboratory evaluations to allow comparability between results from participants.

Laboratory-based evaluation sessions allowed for testing of the efficiency and effectiveness of the PATHS software under controlled conditions, and in-depth feedback to be captured from participants on usability and satisfaction. A user-centred methodology was employed based upon the Interactive Information Retrieval (IR) paradigm and widely adopted in the IR community for more complex systems with functionality going beyond simple search. This Interactive IR evaluation utilises simulated and natural work tasks typical of user information needs, as an aid to system interaction in a controlled environment. In this way it has been possible to evaluate to what extent the system supports the user in their specific context. Extensive data capture via observations of user activity (e.g. screen capture, eye-tracking and transaction logs) was complemented by quantitative and qualitative data about the participants, their information behaviour and their experience of completing the tasks and using the system in general, collected via questionnaires and interviews. The user evaluation was carried out by MDR and USFD.

Evaluation of the prototype system was also conducted by the technical development partners; i-Sieve, Avinet, The University of the Basque Country and The University of Sheffield. These evaluations were concerned with the system architecture, content

processing and enrichment and user interface design and focussed on systematic, objective evaluation of the building blocks of the system as standalone entities. These technical activities informed, complemented the overall user-centred evaluations.

For the final round of evaluation groups of target users were invited to take part in Field Trials of both the main Prototype and the Mobile Application. Data was collected in using a more naturalistic approach than previous evaluations, and provided an opportunity for the collection of qualitative data from focus group and interview discussions and quantitative feedback through questionnaires. The field trials were carried out by MDR and USFD. Again, evaluation of the prototype system was also conducted by the technical development partners; i-Sieve, Avinet, The University of the Basque Country and The University of Sheffield.

5 Intellectual Property and Ownership

This chapter identifies the Intellectual property rights (IPR) associated with PATHS products that are held by the project partners either through: project agreements, organizational policies or through constraints propagated from underlying licenses of libraries and components.

5.1 Project IPR policy

The Management of Intellectual Property is considered one of the keys of the project's success. Our strategy is based on the need to manage both rights for pre-existing resources in addition to resources that are developed, enhance or improved during the project's lifetime. Our strategy is intended to provide a framework for the continuing development and improvement of PATHS results.

IPR of cultural heritage assets provided by partners for use during the life-time of the project will be managed as part of the overall strategy. During the lifetime of the project, the IPR of pathways created by users will be managed under an appropriate Creative Commons licence to enable the sharing and reuse of pathways. The technological development derived from the project will be protected by a GPL license (General Public License) where possible, allowing free utilisation, distribution and modification, and the permanence of this copyright policy for any improvements carried out after the product is developed.

The long-term IPR strategy is based on the continuous improvement and exploitation of PATHS results, and the project will release results under appropriate licences to enable utilisation, distribution and on-going development of products. Pre-existing assets are exempted from this strategy (as their IPR is under pre-existing management arrangements).

The management of the copyright and results aims at facilitating and encouraging collaboration among experts in this field, while providing an appropriate framework for the development and continuous improvement of products and delivery of support services required by industry end-users.

5.2 Partner specific IPR conditions

The following section describes the IPR conditions for each PATHS consortium partner.

5.2.1 University of Sheffield

The University of Sheffield (USFD) was awarded university status in 1905 when it received its royal charter. It is a member of the Russell group of research-intensive UK universities and is world leading in both research and teaching activities. The university actively promotes the exploitation of research output through knowledge transfer. Exploitation strategies listed in section 8 will be explored on a case-by-case basis. Input from USFD's Research and Innovation Services have expertise in commercialisation and knowledge transfer and can advise on any specific constraints. Their input will be sought in connection with any exploitation opportunities.

Note that some of the results from USFD's development efforts in the PATHS project are subject to IPR restrictions propagated from underlying components/libraries licenses as

described above. Note also that some products have been jointly developed with the University of the Basque Country.

5.2.2 University of the Basque Country

The University of the Basque Country (UPV/EHU) is a teaching and research institution officially founded in 1985. The UPV/EHU is at the helm of Research, Development and Innovation in a region that stands out as one of the most prosperous in Europe. The University promotes the exploitation and, where appropriate, the protection of intellectual property in line with one of its missions that is to contribute to the socio-economic development of the Basque region. To this end, different strategies maybe adopted on a case-by-case basis, and there is no restriction for any of the exploitation strategies described in chapter 8.

Note that some of the results from UPV/EHU's development efforts in the PATHS project are subject to IPR restrictions propagated from underlying components/libraries licenses as described above. Note also that some products have been jointly developed with USFD.

5.2.3 i-sieve technologies

For the sentiment analysis software, i-sieve technologies is using technology developed at the National Centre for Scientific Research in Athens. The processing is done from i-sieve's server park in Athens, and the results are made available to clients through a web application where they can monitor the present sentiment towards their specific interest. This is proprietary technology that was brought into the project in a near-complete state and is exempted from the overall GPL licensing policy of the project agreement.

5.2.4 AVINET

The main business of AVINET is the sale of system development services and implementation of web applications within the domain of GIS. The company makes continuous investments into research and development, on average about 10% of the annual turnover is dedicated to R&D activities and innovation.

The company's business model is centred on services and to a lesser degree software licenses. AVINETs software products are built using Open Source components and libraries; the software itself is distributed under a modified L-GPL license that restricts commercial re-use and limits the company's liability in the event of malfunction caused by users' customization - but that permits users to freely modify the work as they require to suit their purposes.

Results from AVINETs development efforts in the PATHS project are not subject to any IPR restrictions from existing policies or propagated from underlying components/libraries licenses. As such, all exploitation strategies described in chapter 8 are possible.

5.2.5 MDR Partners

MDR is a specialist SME which offers services including research and strategic development, consultancy, training, project management and implementation, service development and technical support. It works with libraries, archives, museums, information providers, and other organisations in the cultural heritage, public data and education sectors at European level and internationally.

Results from MDR's user requirements gathering and evaluation efforts in PATHS are not subject to any IPR restrictions.

5.3 Summary of IPR Constraints

The IPR associated scientific research and knowledge products are made clear for each publication and/or project. All results are made public in accordance with the project IPR policy.

The IPR associated with the methodologies used to execute manual services within the context of the PATHS project have also been clarified. The methodologies are described in detail in public deliverables and may be freely reused by third parties seeking to provide services including:

1. content enrichment
2. user testing, user trials
3. user requirements specifications

With knowledge and services catered for, it is necessary to determine the IPR conditions for each of the potential software products that can be brought to market in follow-up of the PATHS project. The below table summarizes the conditions for each product and technical partner.

The first column lists the name of software products; the second column identifies the license under which the software is available for exploitation and the third column identifies who is the owner of intellectual property associated with each of the products.

The term 'IP owner' signifies the entity that was the lead developer(s) and originator of the software within the PATHS project. This does however not restrict other partners, groups of partners or third parties to exploit the product as long as the terms and conditions of the specific license are observed.

MDR Partners and Alinari (the latter party ceased) have no stake in software development or distribution and have therefore not been included in the matrix below. They do however enjoy the same rights as other parties in agreement with the respective licenses.

Product	**License**	**IP owner**
NLP tools	GPL + propagated terms & conditions from underlying licenses	USFD, EHU
Sentiment analysis tools	Proprietary, exclusive	i-sieve technologies
Web API and information retrieval	GPL	USFD, AVINET
Recommender system	GPL	USFD, EHU
"Map-based" visualization of thesauri	GPL	USFD

6 Market Study

6.1 Factors in defining the market

As a whole, the PATHS technology suite is targeted at actors within the Europeana family of content providers and aggregators. In order to reach out to a selection of actors that have the potential to reflect upon the entire suite of products, this is the source from which we compiled our sample.

We have targeted three different types of actors:

1. Content providers: institutions who have their own collections of content and publishes this both directly to the Internet through its own web infrastructure as well as via the Europeana aggregation infrastructure

2. Aggregators: institutions who are aggregating content on either regional or national level

3. Technology providers: SMEs within fields such as software development, software customization and data processing who are offering commercial consultancy services to actors in the two first groups.

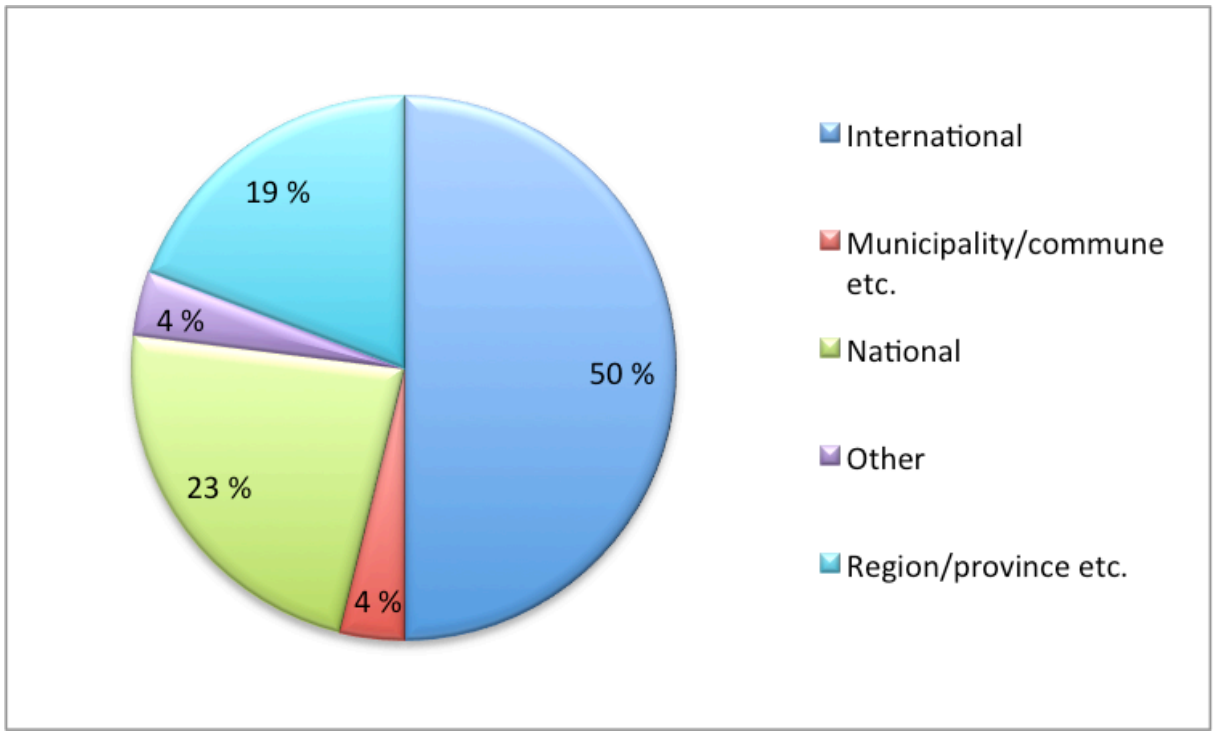
The latter group is particularly interesting as PATHS software products to some extent are modules and libraries rather than complete solutions. Whereas cultural heritage professionals mainly would adopt turnkey solutions, technology providers are more likely to adopt GPL licensed technology into their own products, securing a wider uptake and dissemination of PATHS products than what would otherwise have been possible.

When split into its individual research, software and service products, PATHS results have a potentially much wider range of applications that are not confined to the cultural heritage domain. The market study therefore sheds light on only one of a range of thematic domains where some of the products may be applied.

6.2 Sampling

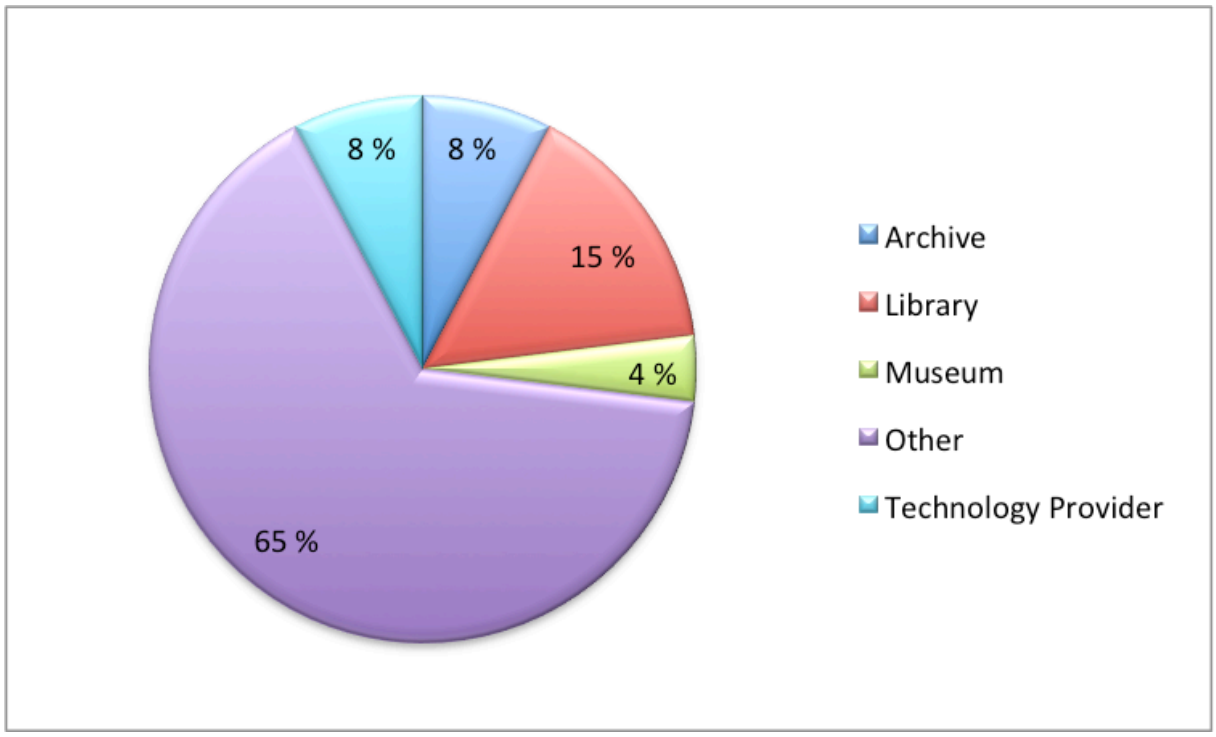
In order to identify the high-level attitude towards PATHS products and services in the defined target market, a *judgement sample* was prepared by AVINET and MDR by selecting leading representatives of key Europeana related activities for the past five years

Figure 2: Breakdown of geographical service area of the respondents



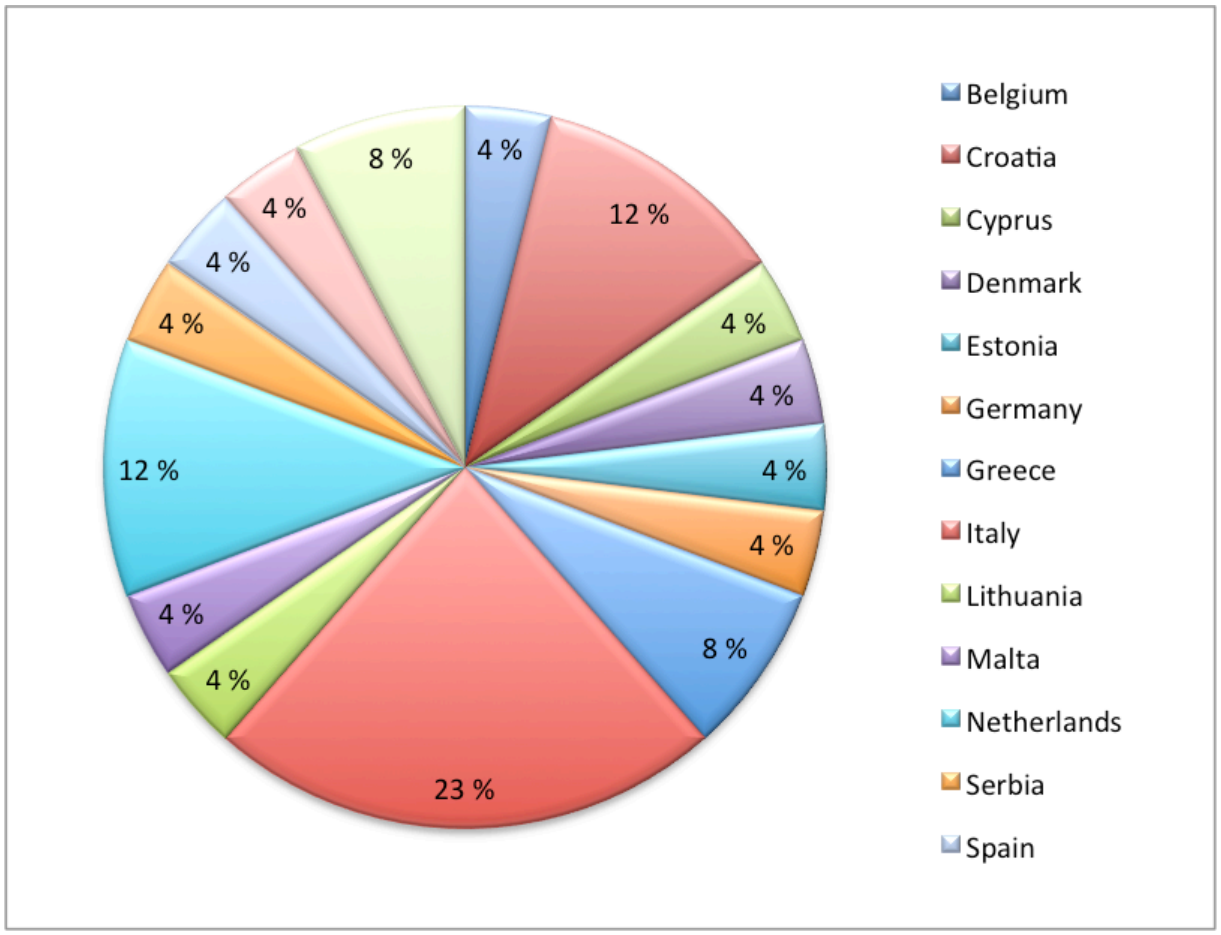
As the diagram above shows the international and national level is overrepresented in the survey. The local level, on municipality or commune, is only represented by 4 % of the responders. This aspect is important to remember when interpreting the results from the study.

Figure 3: Breakdown of organization types



The “other” category includes universities, think tanks, government ministries, SMEs, technology providers and research institutions.

Figure 4: Where do the recipients come from?



Thirteen different countries are represented in the sample. The sample represents the majority of Europeana partners and provides a broad picture of the perception of PATHS products in the potential market.

6.3 Execution

The survey was implemented using the free, open source tool LimeSurvey. The survey was distributed to 170 selected participants, described under the heading “sample” above, by e-mail. After two weeks, a reminder was issued to all invitees that had not either opted out of the survey or already completed the questionnaire.

The survey was active for three weeks from the 31st of October until the 23rd of November 2013 - in all three weeks.

6.4 Results of Market Study

The market study received a limited number of responses. Out of the 170 selected recipients who were invited to respond, a total of 26 completed the survey, 16 started but did not complete the survey and 14 opted out of participating.

This gives a response rate of 33%. The information retrieved from the study is therefore considered qualitative rather than quantitative and most trends will require additional market probes prior to rollout.

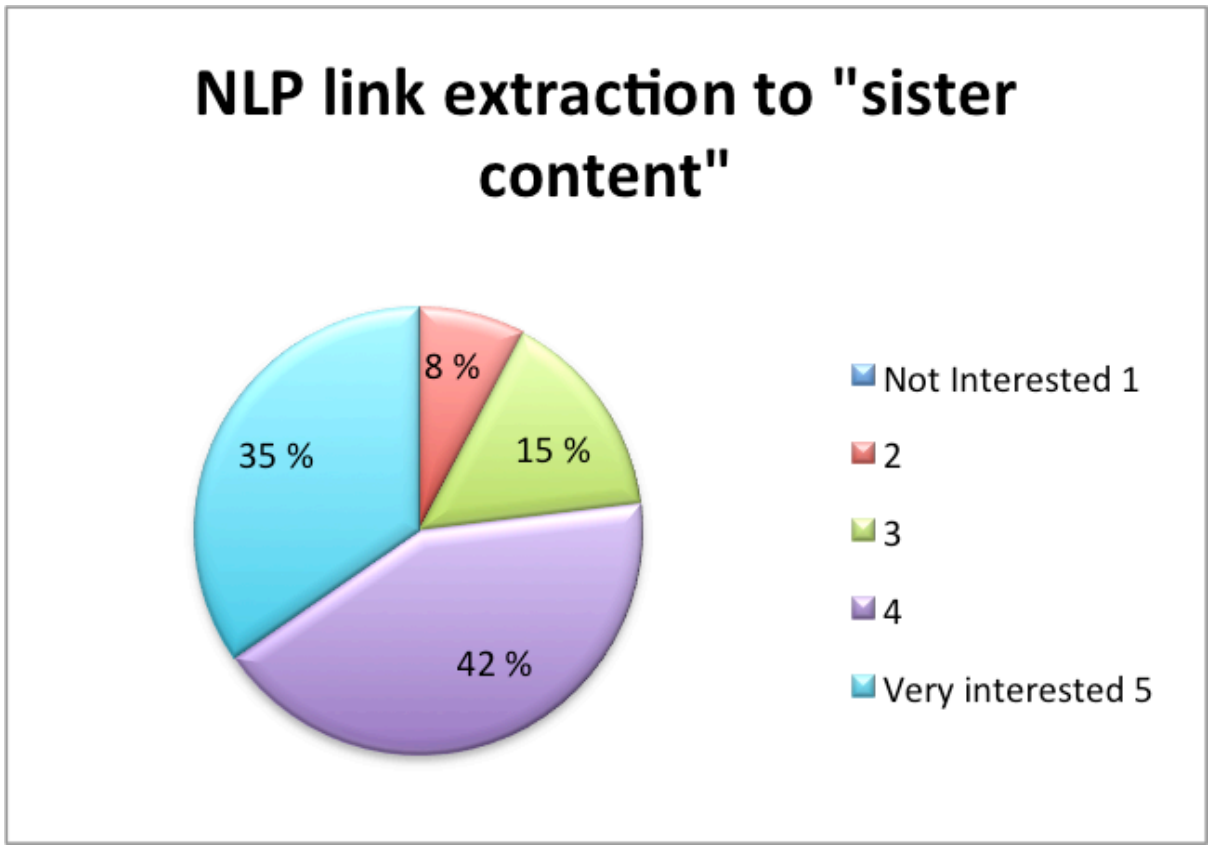
However, some patterns and trends were nearly unanimous across all the responses and can be interpreted as strong indications, e.g. attitudes to Open Source software.

6.4.1 Attitude to Paths Technologies

In the following a scale from 1 to 5 is used where 1 is not interested and 5 is very interested. The recipients are asked to rate the technologies according to how interesting they find them for their institution.

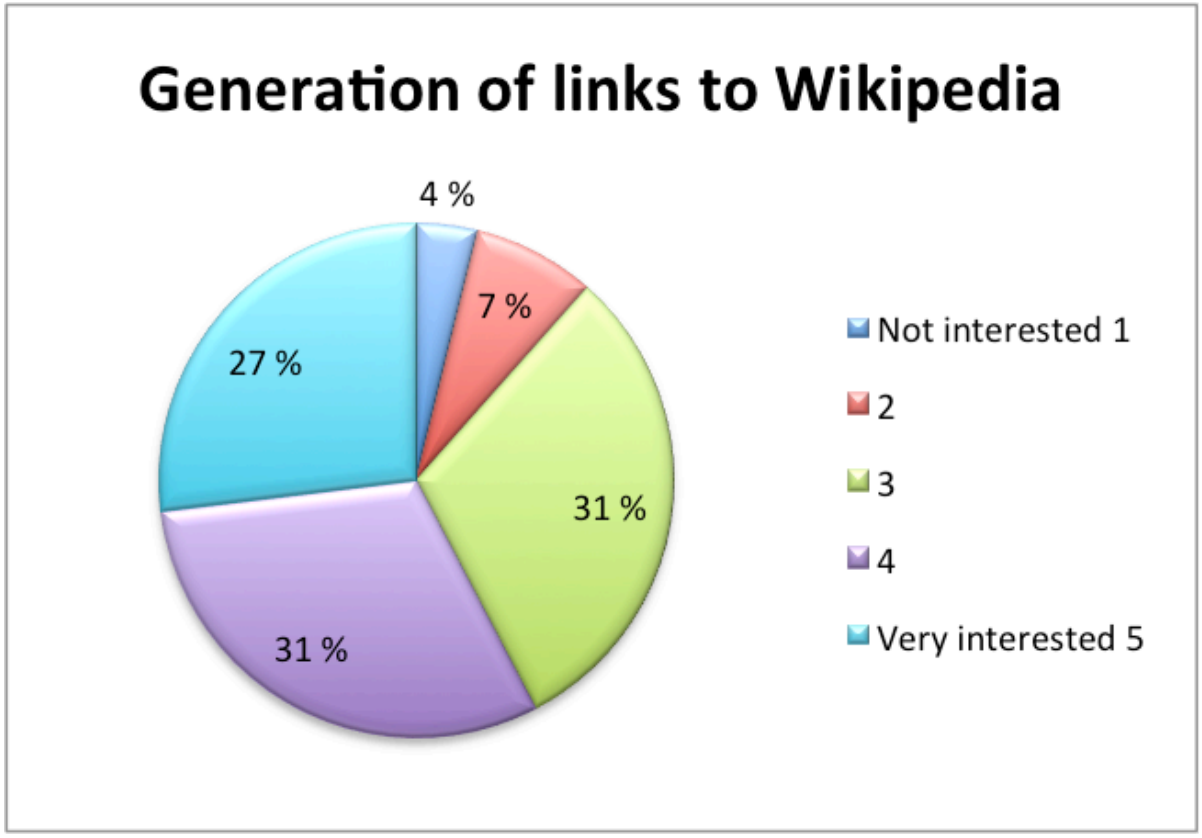
6.4.1.1 Natural language processing

Figure 5: NLP link extraction to "sister content"



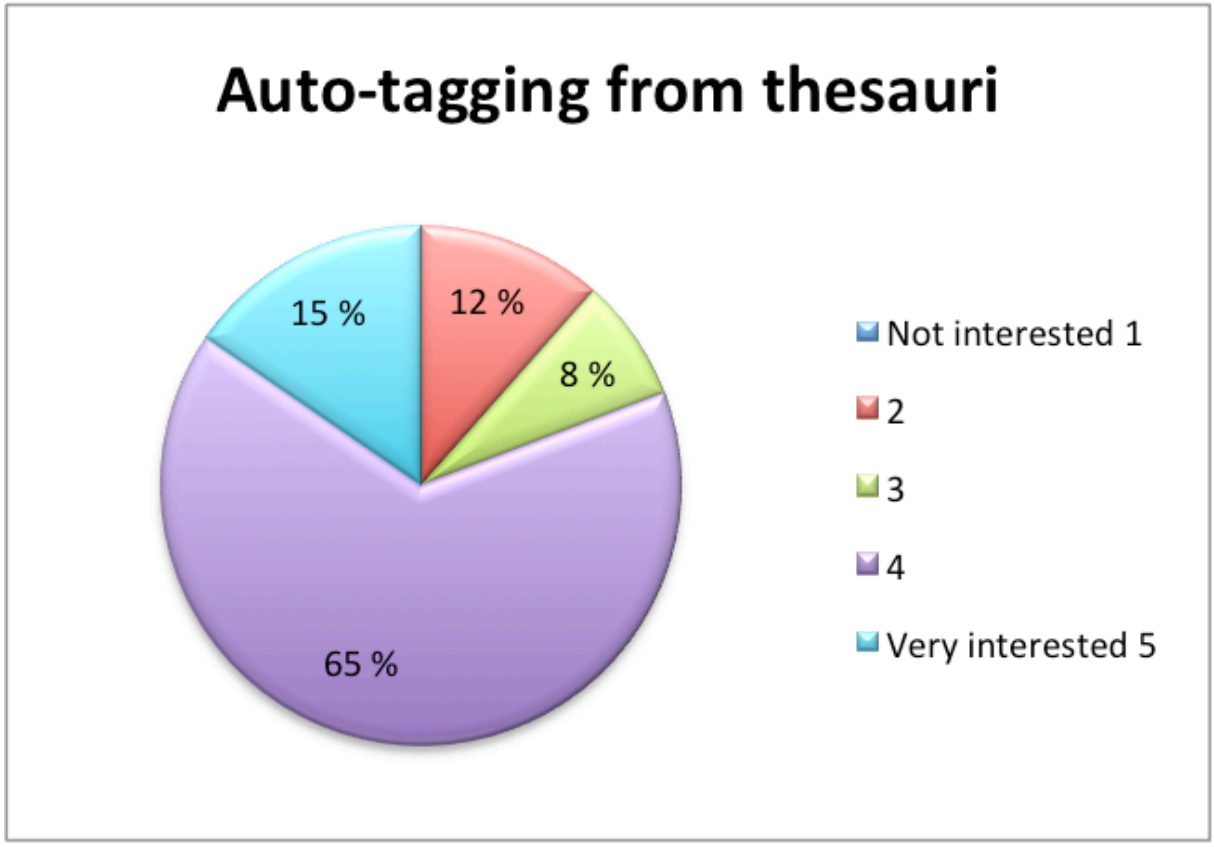
The respondent are positive to link extraction to "Sister content" and a total of 77 % reports that they are interested or very interested (score 4 or 5).

Figure 6: NLP link extraction to Wikipedia



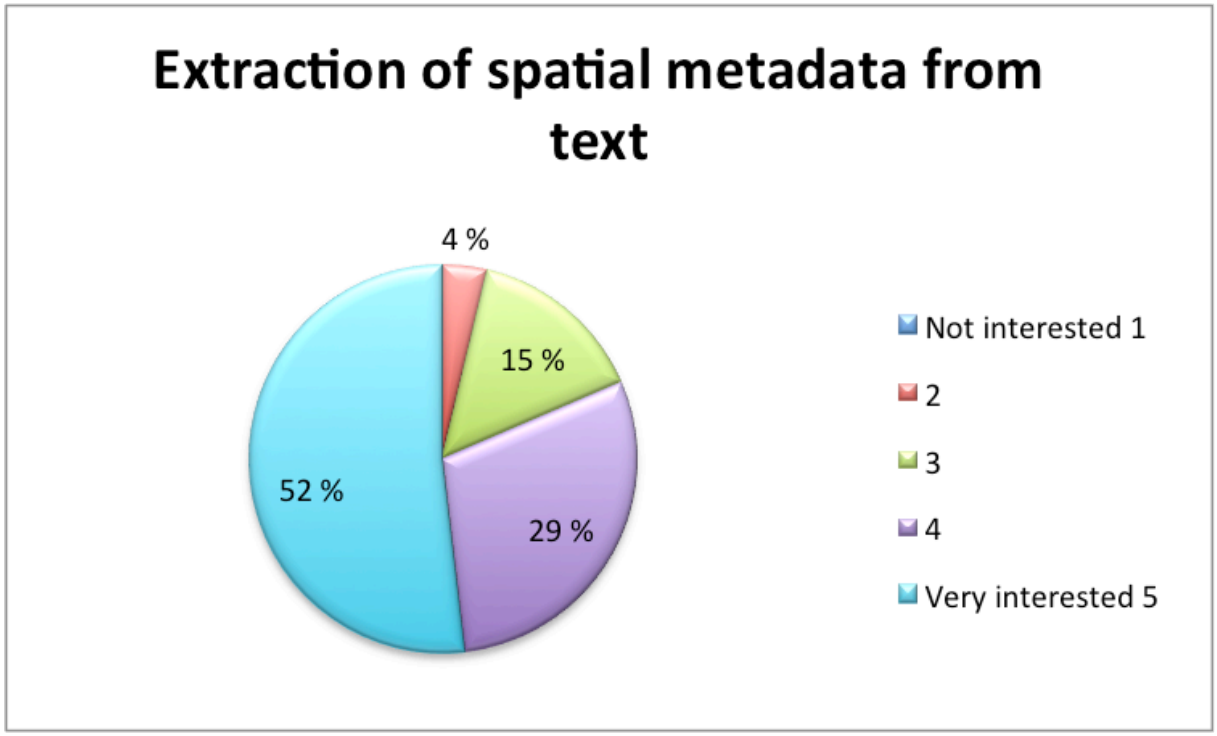
The extraction of links to Wikipedia is also considered as interesting, but not as interesting as the “sister content”. A total of 58% reported a score of 4 or 5.

Figure 7: NLP auto-tagging content with terms from vocabulary



Auto-tagging content with terms from vocabularies are regarded as interesting to 65% of the respondents and very interesting to 15%. This may indicate that there is a possible market potential for this kind of knowledge/software/service.

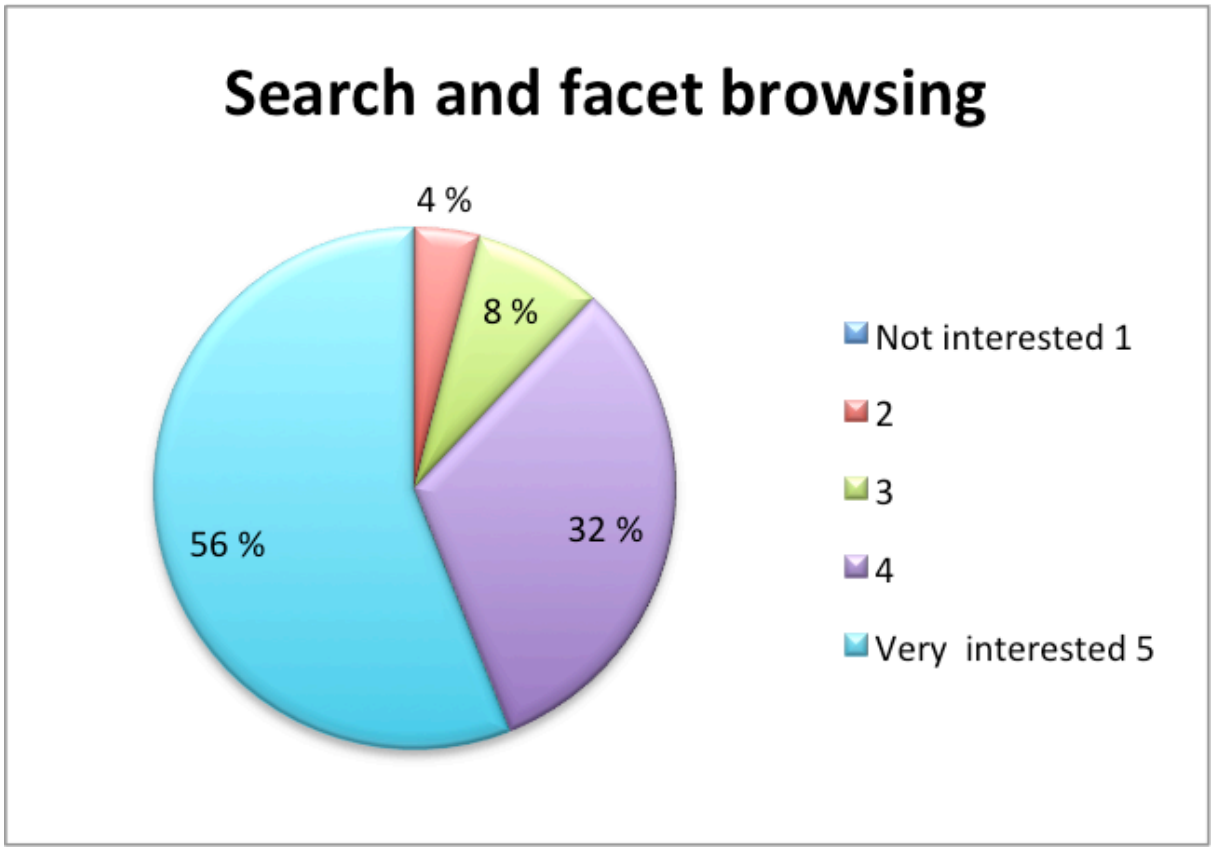
Figure 8: Extraction of spatial metadata from text



Extraction of spatial metadata from text also have considerable possible market potential. 52% of the respondees say that they are very interested (score 5) in this. Another 29% are interested (score 4).

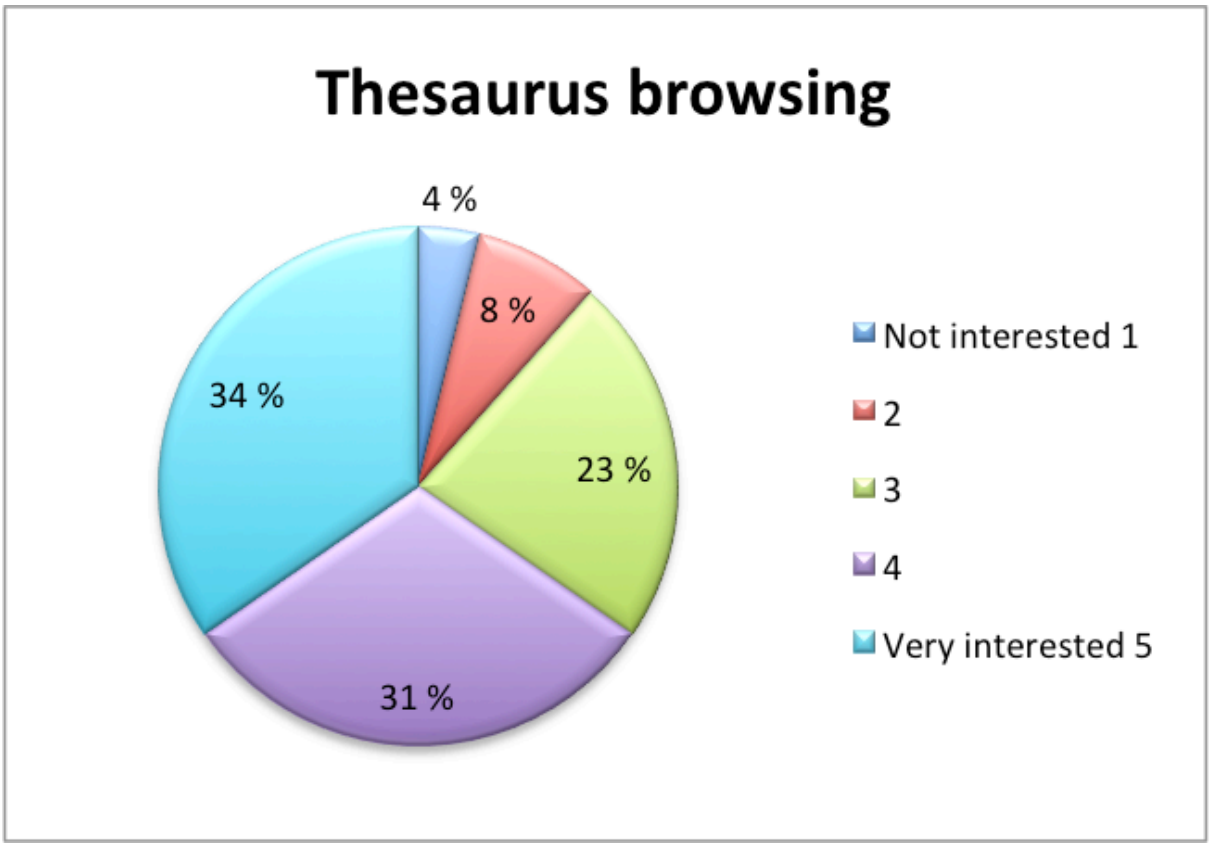
6.4.1.2 Information retrieval

Figure 9: Search and facet browsing



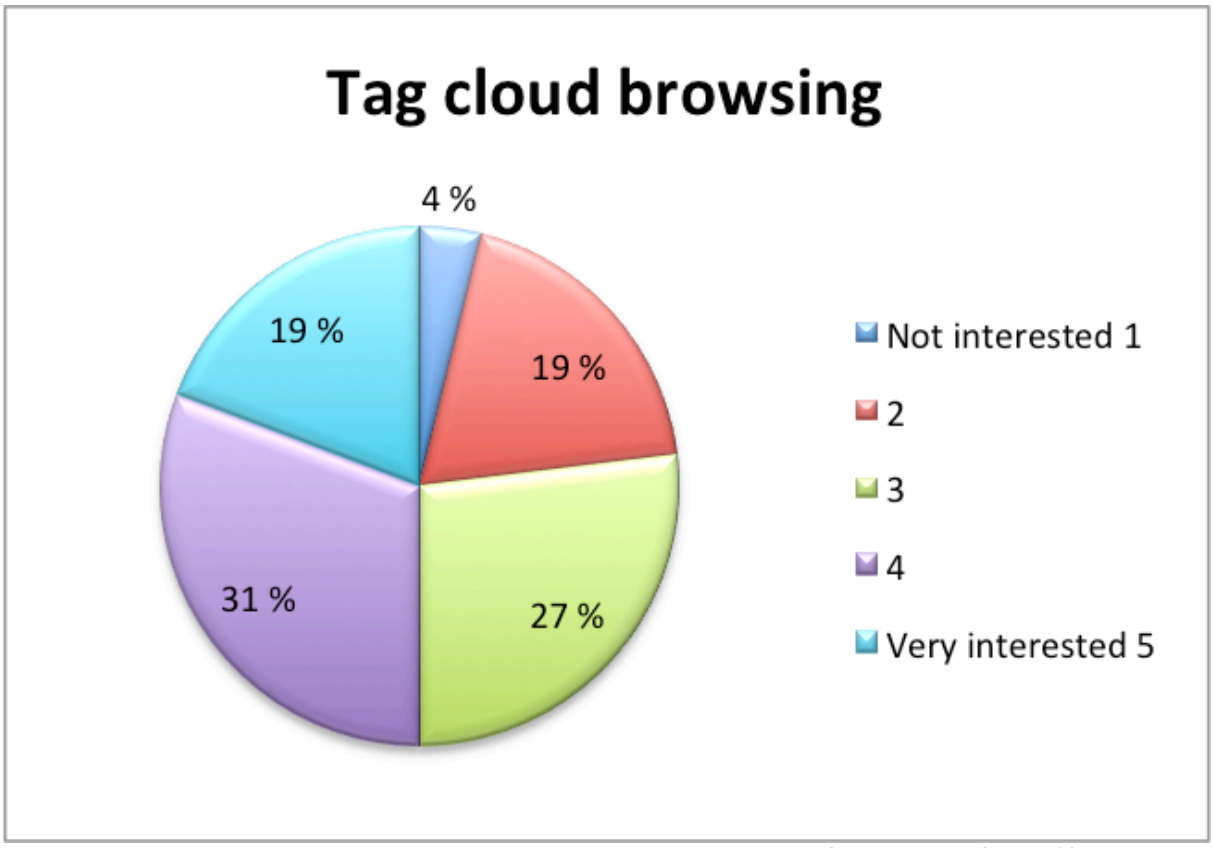
88 % of the total are interested or very interested (score 4 and 5) in search and facet browsing. This is the highest score on the survey and we consider this knowledge to have a high potential in the market.

Figure 10: Thesaurus browsing



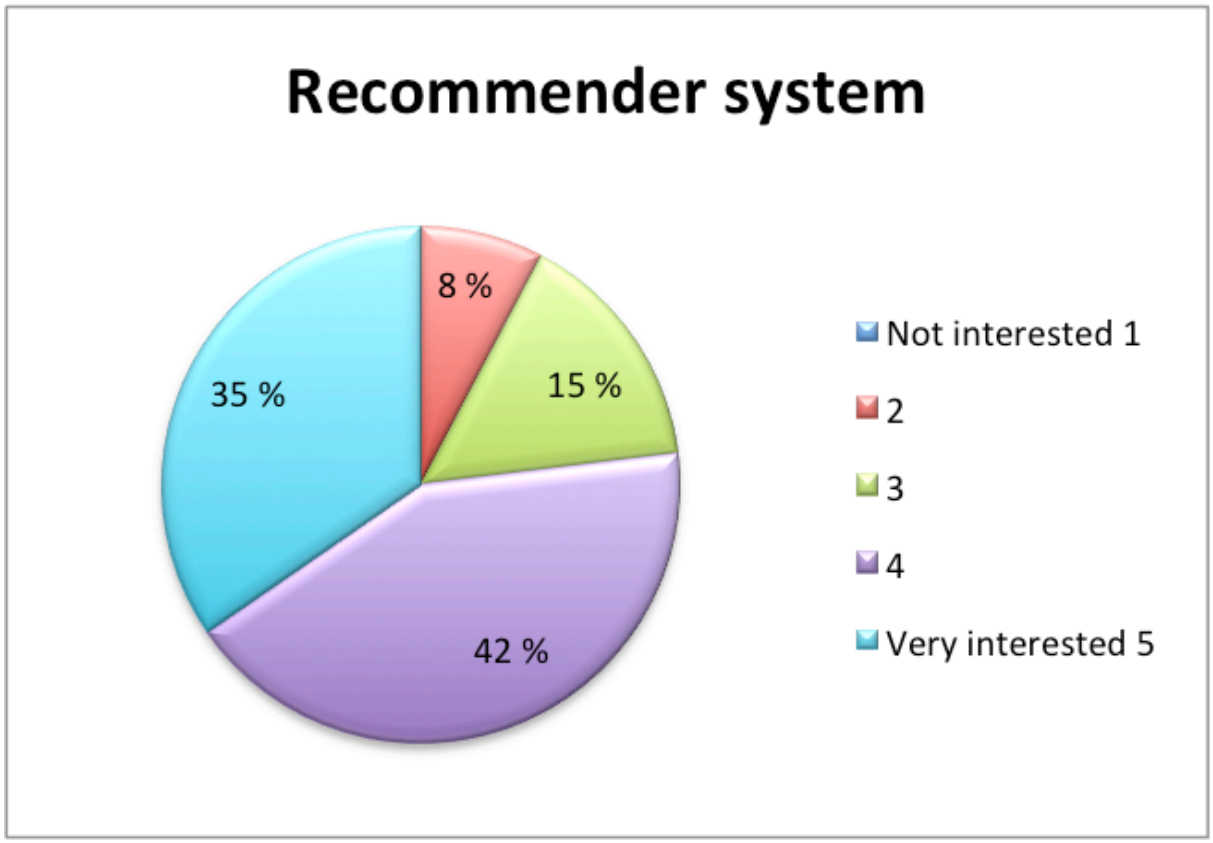
Information retrieval through thesaurus browsing ends up with 65 % score on either interested or very interested (score 4 and 5). This indicates that within information retrieval thesaurus browsing is not as interesting as search and facet browsing for the respondents.

Figure 11: Tag-cloud browsing



Tag cloud browsing was less popular among the respondents. Only a total for 50% says that they are interested or very interested in this way of information retrieval.

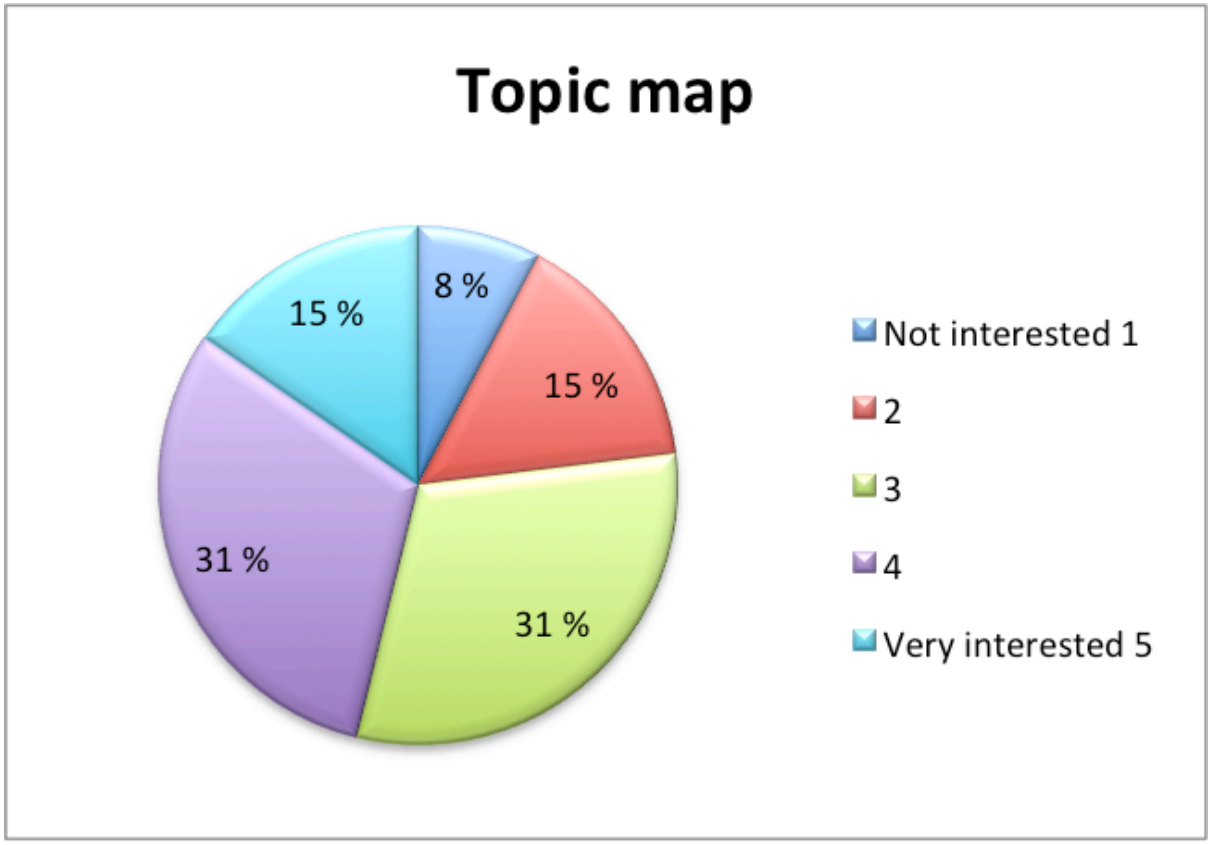
6.4.1.3 Recommender systems



Recommender system seems to be popular among the respondents. 77% says that they are interested or very interested in this system. Software products based on this technology have good market potential.

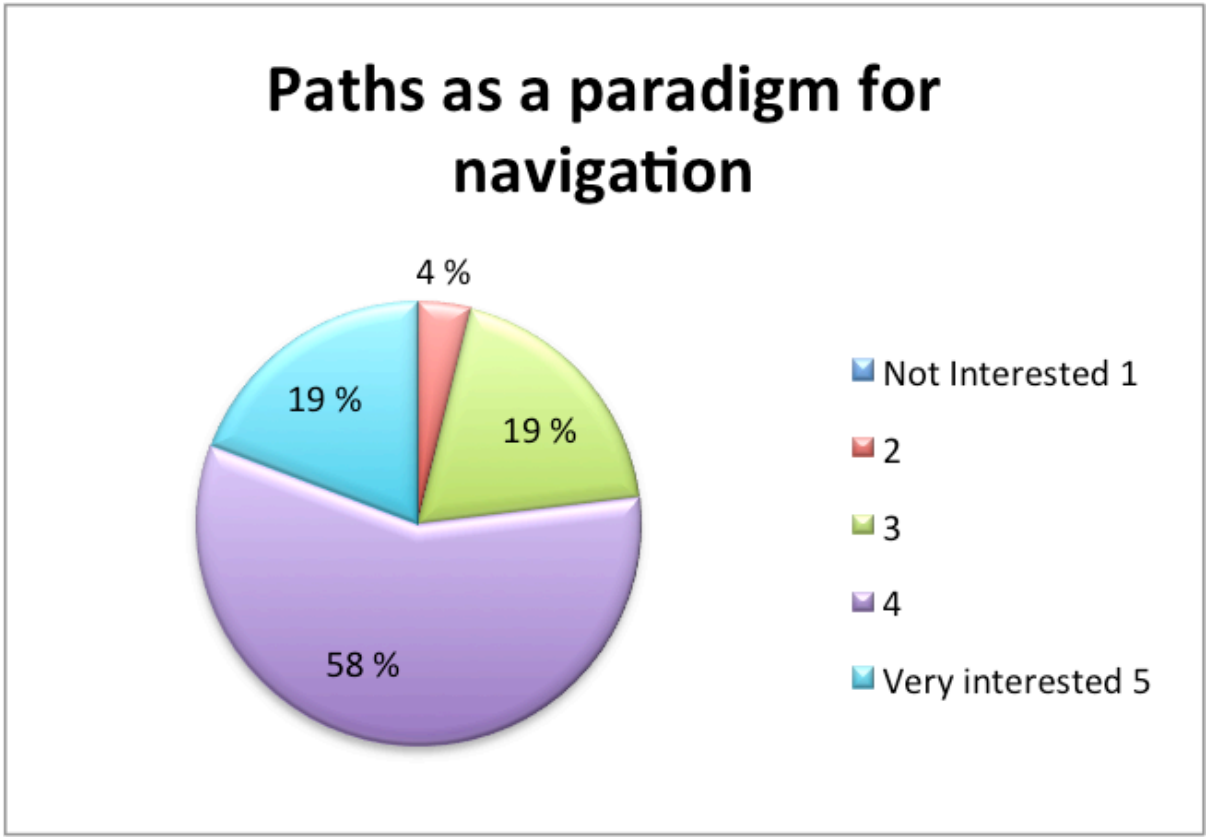
6.4.1.4 Advanced information visualization

Figure 12: Topic map



Visualisation of information through topic maps were not considered to be very interesting among the respondents. Only 46% gives this technology score 4 or 5, and as much as 8% are not interested at all.

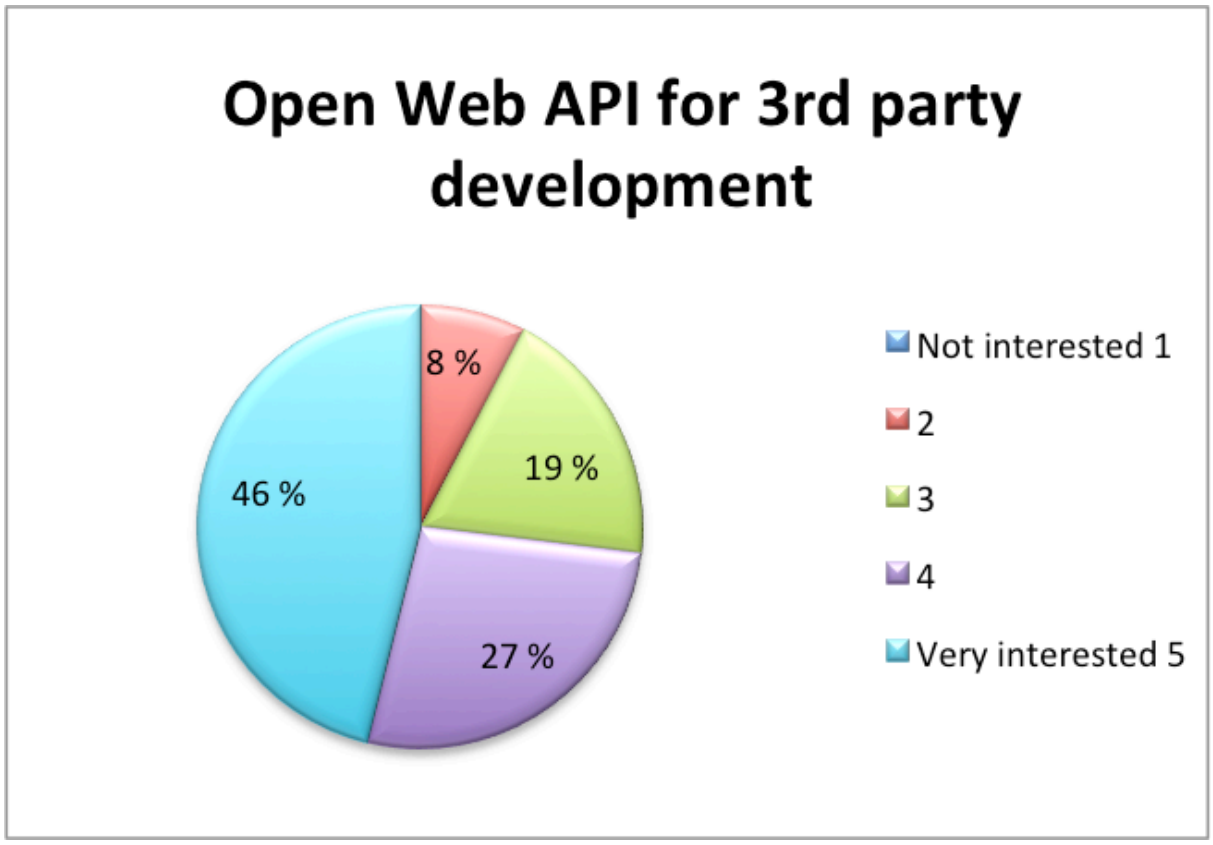
Figure 13: PATHS as a paradigm for navigation



Using Paths as a paradigm for navigation is more popular. A total of 77 % says that they are interested or very interested (score 4 and 5) in this way of navigation.

6.4.1.5 Distributed systems

Figure 14: Open Web APIs for 3rd party application development



Using an open Web API for 3rd party development is quite popular. 73% responded with a score of 4 or 5. There may be a potential market for this kind of technology based of the market survey.

6.4.2 How the Market Buys Software

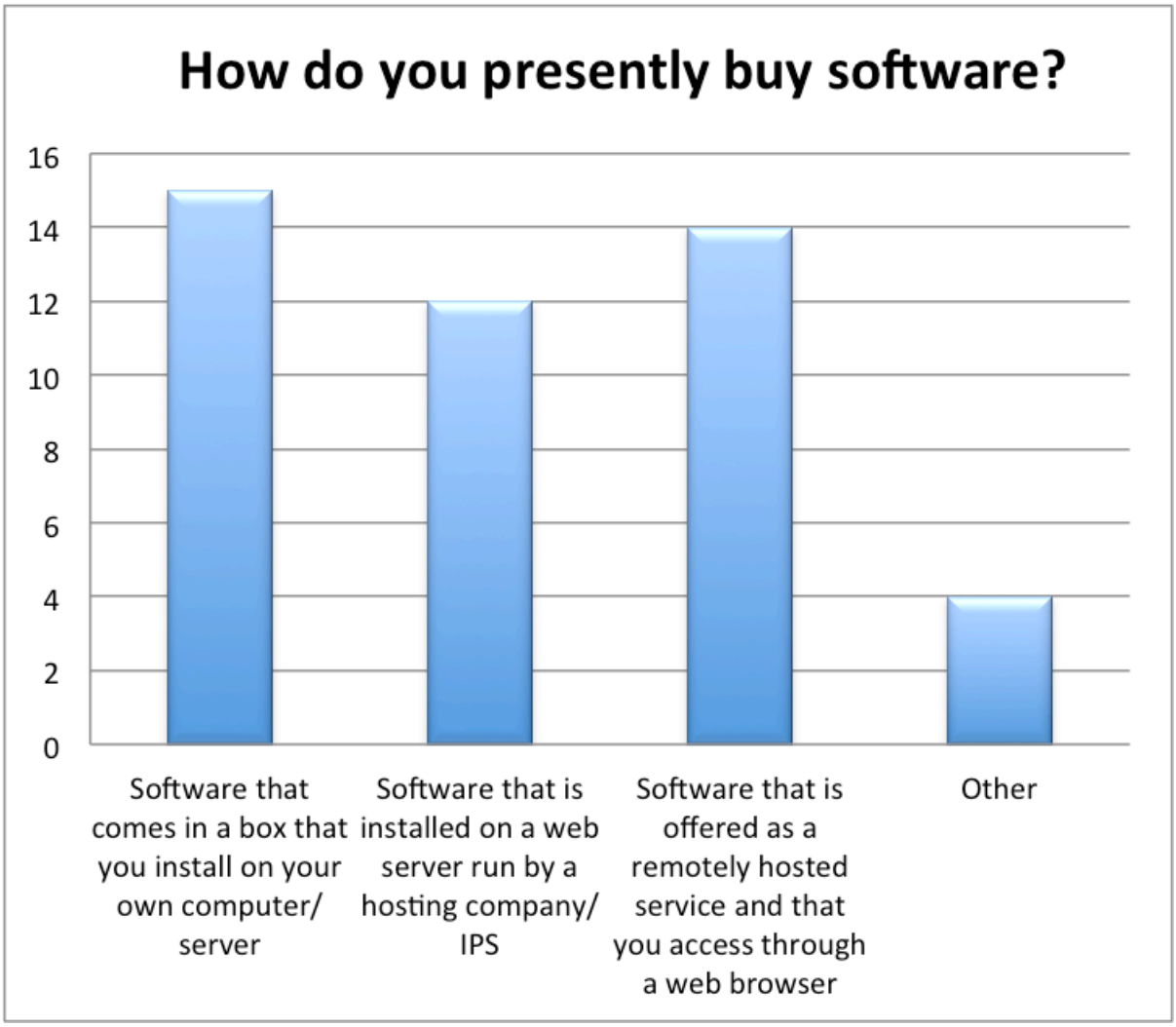


Figure 15: How to by software

Still software in a box seems to be the most popular way to by software, but web-based alternatives also seems to be well known and used.

6.4.3 Attitude to Distribution via the Cloud

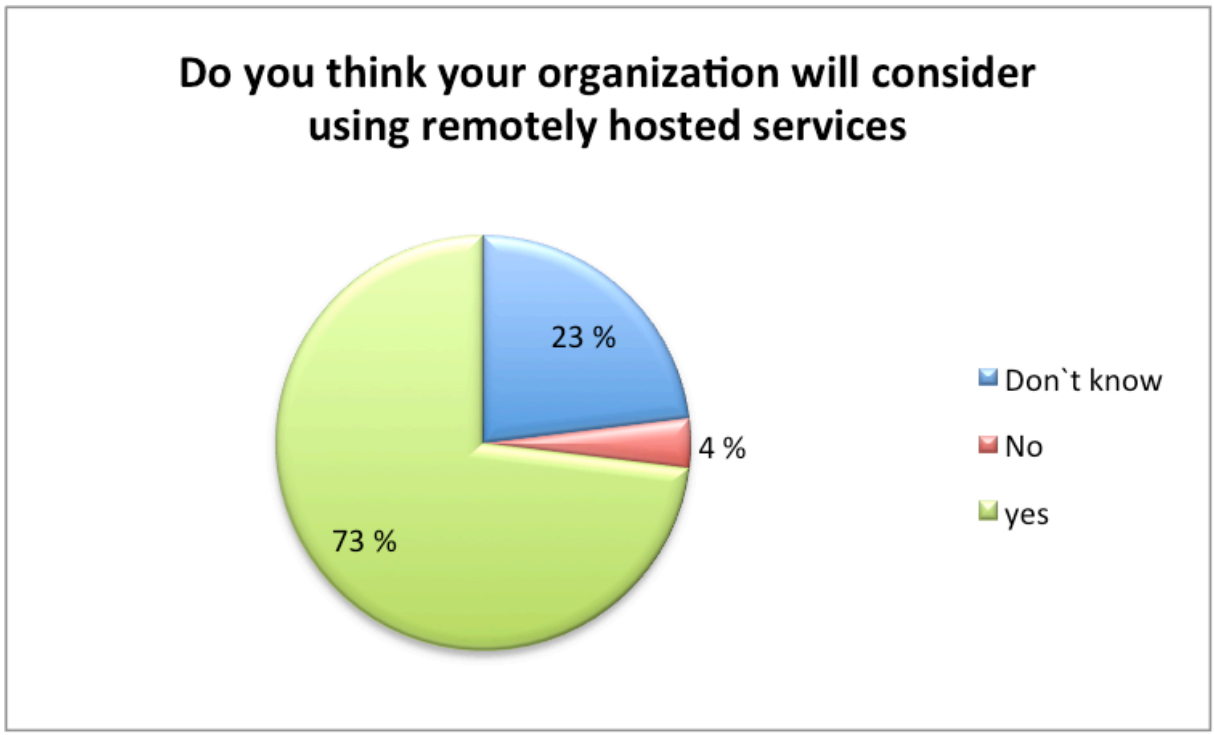


Figure 16: Attitude to the cloud

There doesn´t seems to be any scepticism against remotely hosted services. Only 4% reported that they don´t think their organization will consider using remotely hosted services. This indicates that the Cloud is the way to distribute software and services to the market.

6.4.4 Types of Technology Services Procured

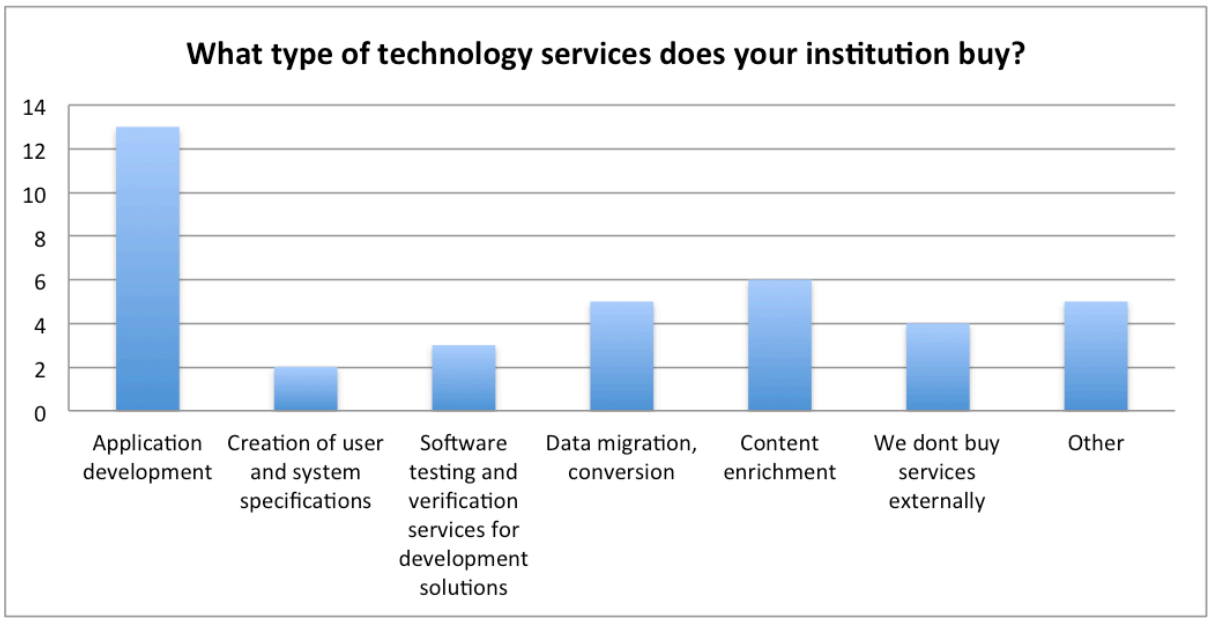


Figure 17: Services procured

When our respondents' institutions buy technology services the most common type is application development. This may indicate that only few of the institutions have this kind of competence in-house and therefore have to buy this when they need it. There could be a market for application development in the GLAM sector.

6.4.5 Established Channels for Marketing

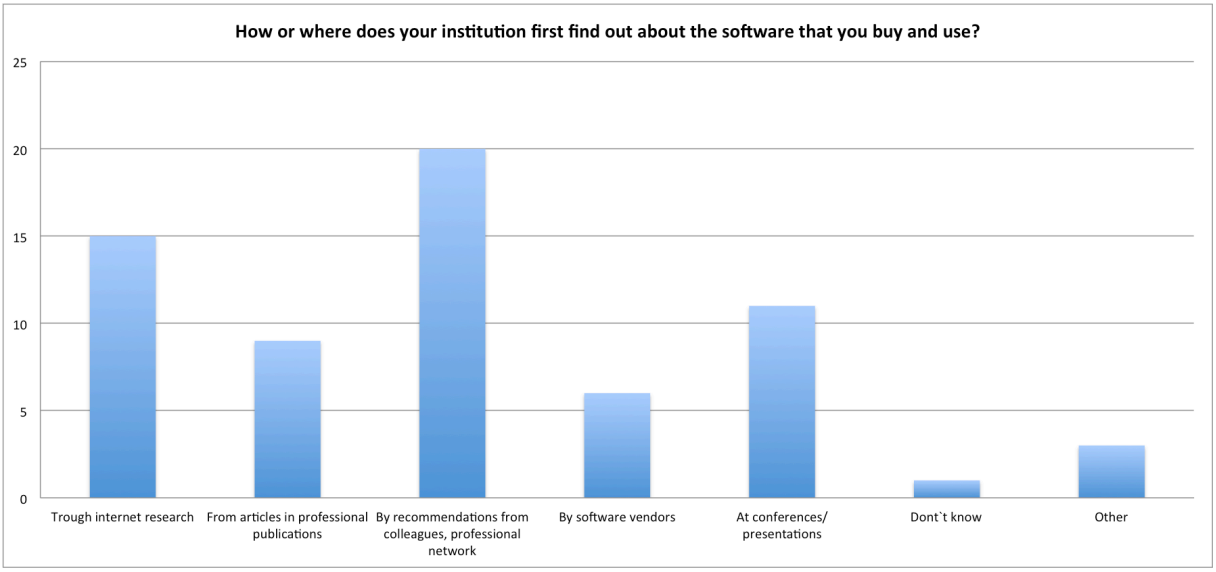
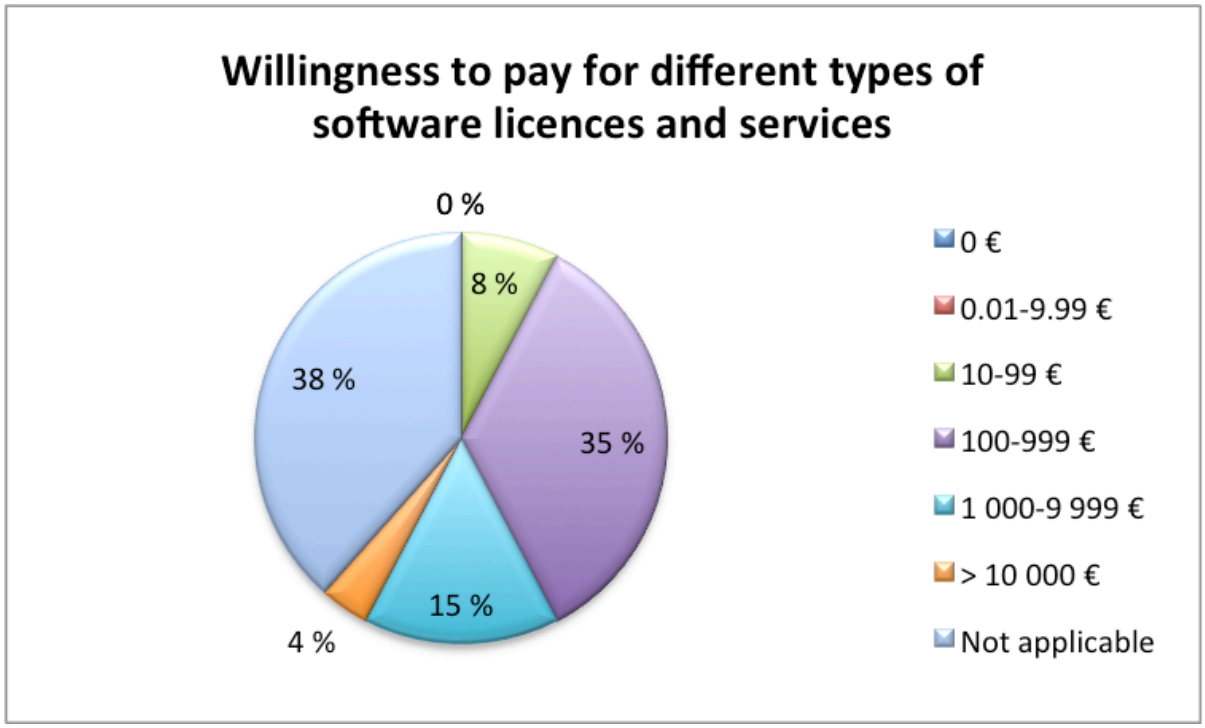


Figure 18: Channels for marketing

There appears to be a strong professional network within the GLAM sector and that recommendations by colleagues seems to be the most important way to find out about new software and services. Also Internet research is a common way of getting information about relevant products.

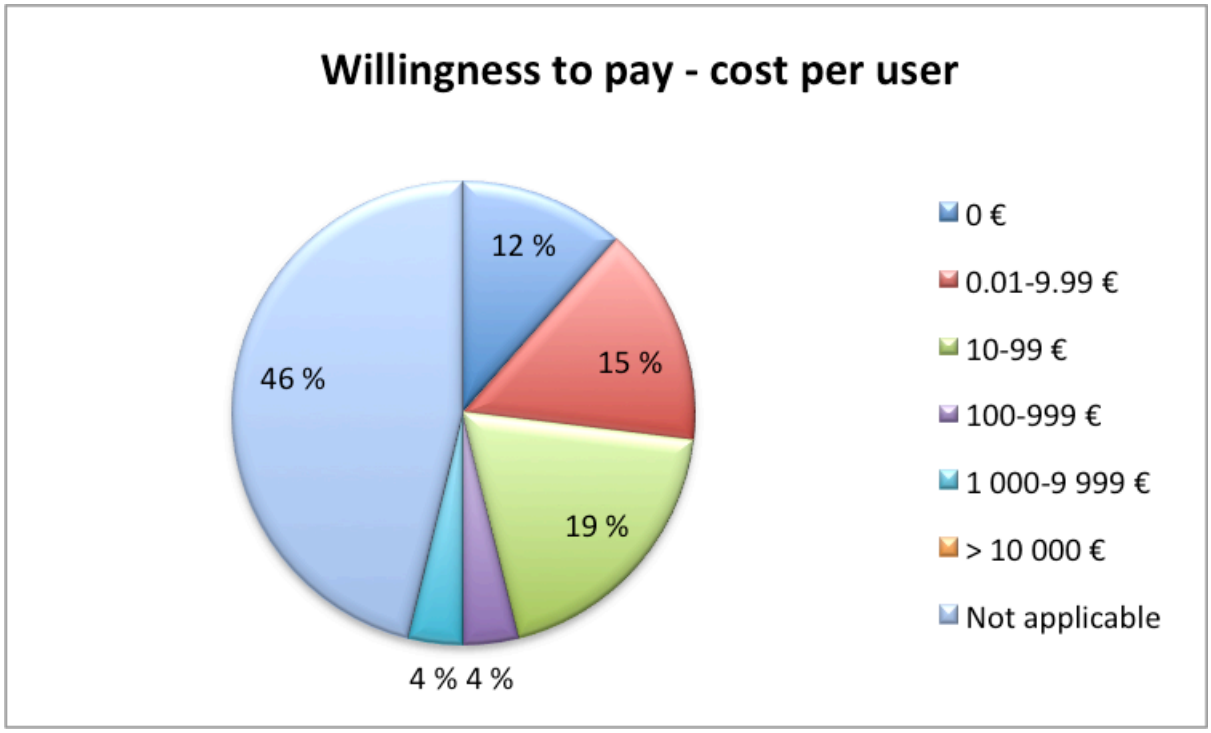
6.4.6 Business Models in Use

Figure 19: Willingness to pay - one time purchase



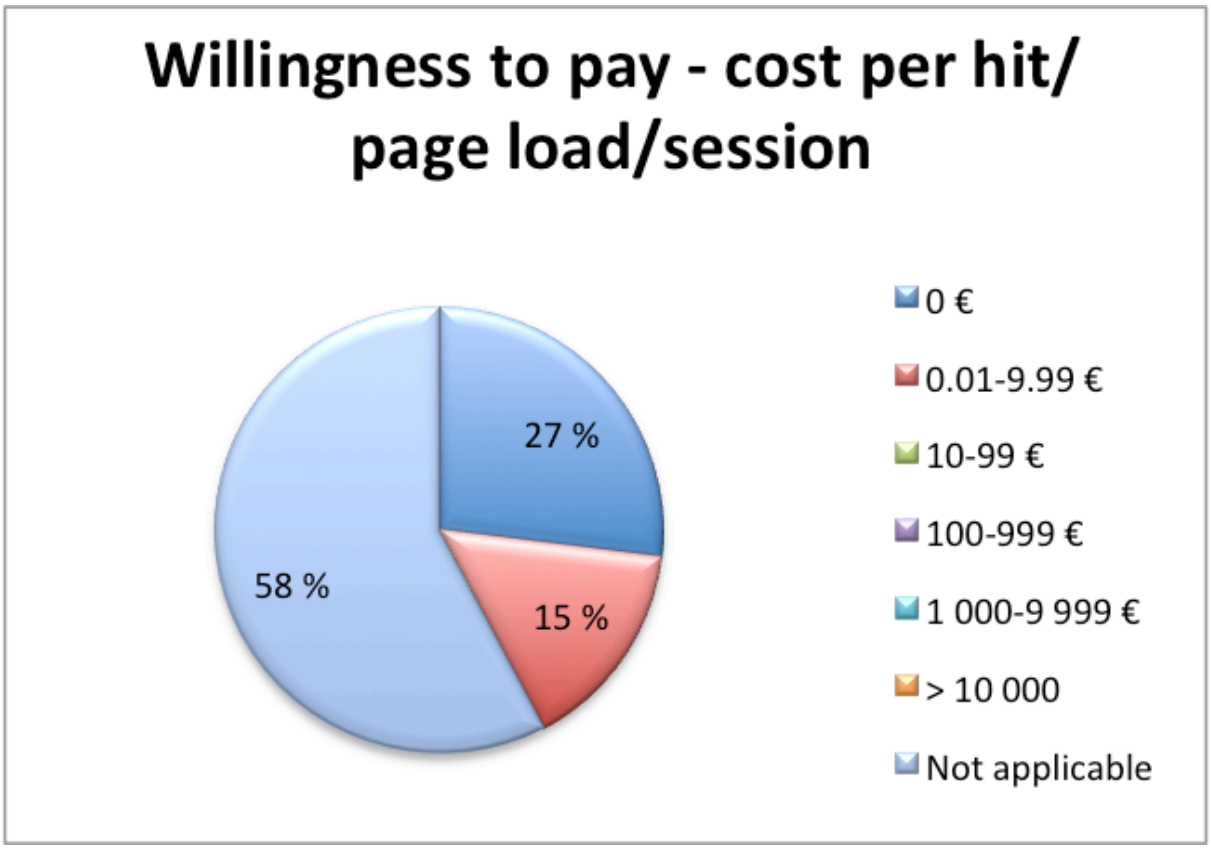
For one time purchase the willingness to pay for software licences and services lies between 100 and 10,000 €. Some are even winning to pay more than 10,000 €. But it is worthwhile to note that as much as 38% of the respondents says that this business model is not applicable.

Figure 20: Willingness to pay - cost per user



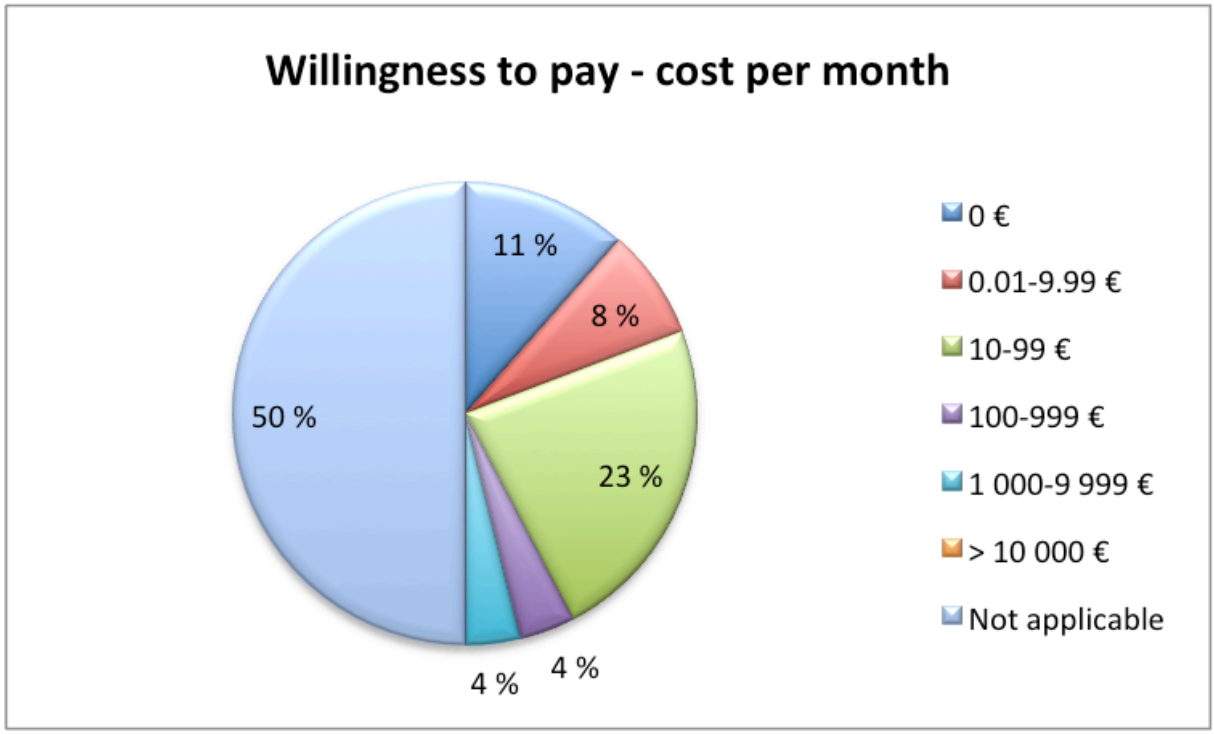
When it comes to cost per user the willingness to pay is quite low and also 46% thinks that this is not applicable. 15% are willing to pay 0.01 - 9.99 € and 19% are willing to pay up to 99 €. But as much as 12% are not willing to pay at all.

Figure 21: Willigness to pay - cost per hit/page load/session



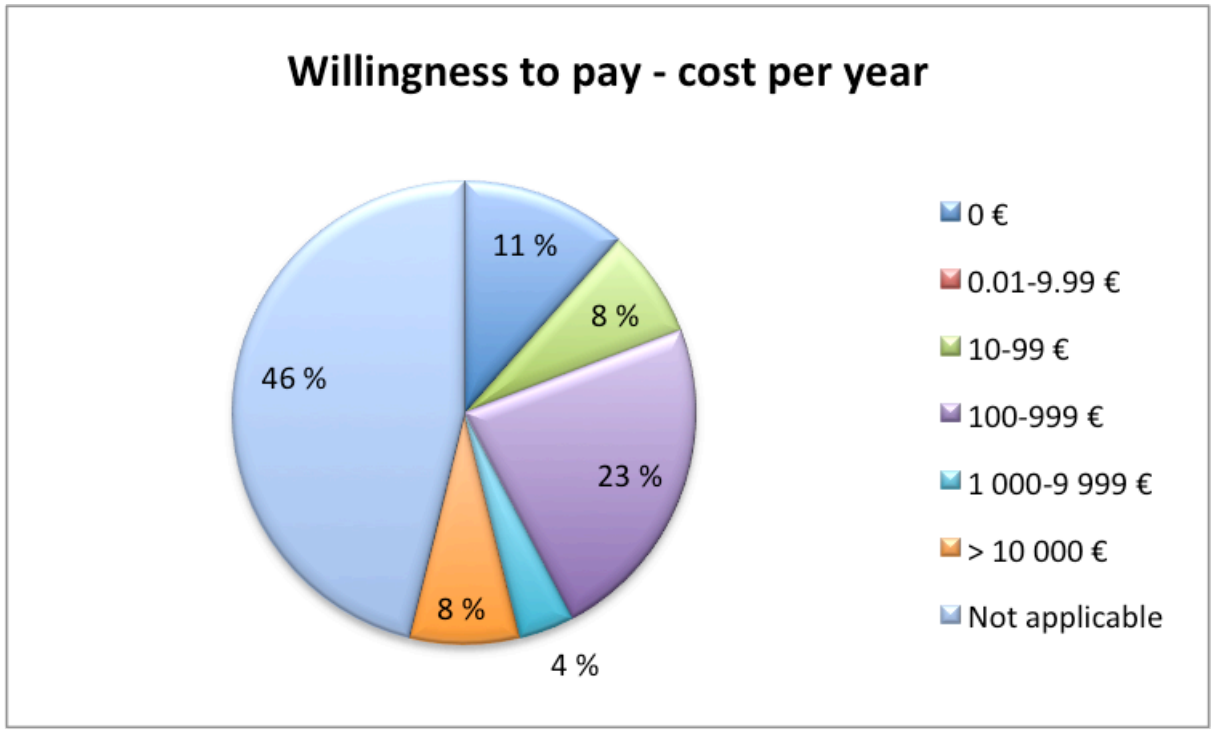
Most of the respondents are not willing to pay per hit, page load or session or find this business model not applicable. Only 15 % are willing to use this as a business model and are willing to pay 0.01-9.99 €.

Figure 22: Willingness to pay - cost per month



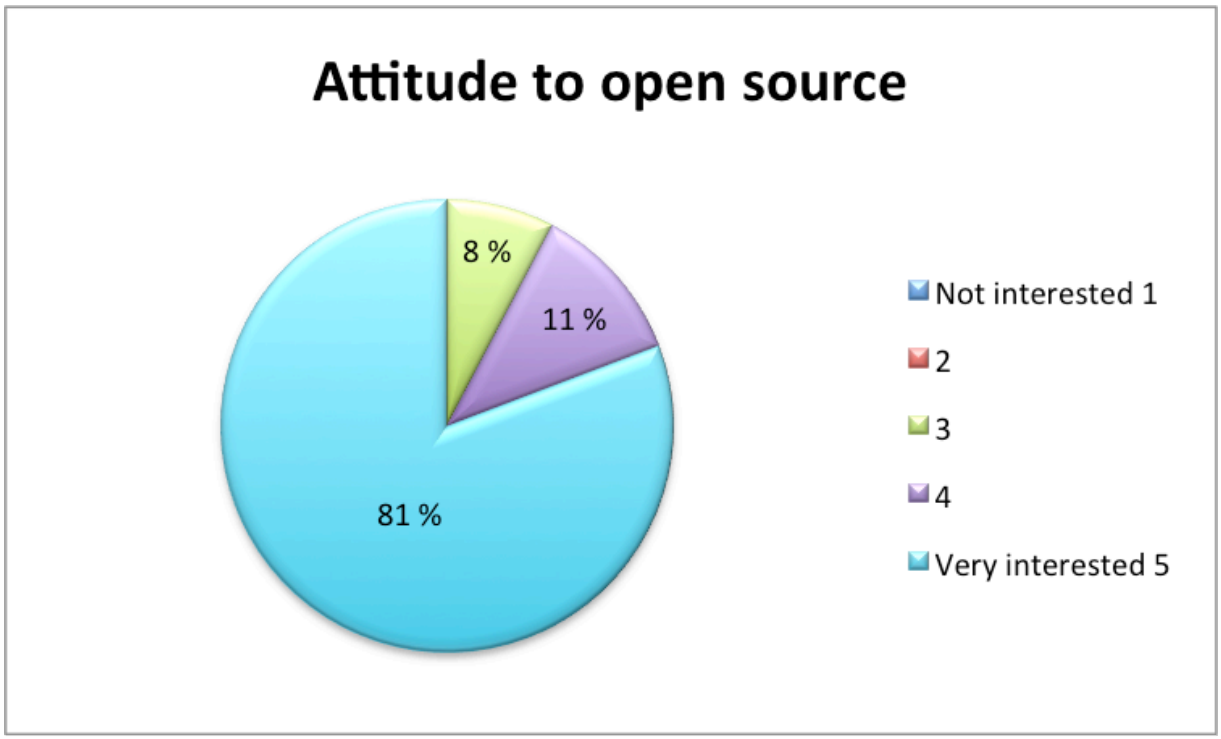
Paying per month does not seem like a relevant business model. 50 % of the respondents answered not applicable. Among those willing to pay 10-99 € is the most acceptable price.

Figure 23: Willingness to pay - cost per year



Paying per year also seems to be little relevant, with 46% using the Not applicable option. Nevertheless, 23 % are willing to pay 100-999 € and 8% are willing to pay more than 10,000 € per year.

6.4.7 Attitude to Open Source



One of the most convincing results in then survey is on the question about attitude to Open Source. 81 % answered that they were very interested in Open Source and none says that they not were interested. This is a strong indication on that software and services based on Open Source components will be marketable in the GLAM sector.

7 Market Potential

To assess the market potential of the PATHS products and services we carried out a SWOT analysis. In this matrix the strengths, weaknesses, opportunities and threats are evaluated. This, combined with the finding from the market analysis, gives an indication on which of the products and services that have a greater market potential.

The SWOT analysis (figure 24 below), together with the market analysis, show that all new knowledge taken forward in the PATHS project are interesting for the potential market. But some limitations appear and there are a lot of possible competitors in the same market as is the case for using technologies from PATHS as a paradigm for navigating cultural heritage collections. Others are also working on similar navigation tools.

Natural language processing tools also have a potential market, but the technology is rapidly changing and can be superseded by other technologies and initiatives.

When asking the potential market the response is most positive towards information retrieval technologies and in particular search and facet browsing. Further development and research in this field may be give the most marketable knowledge and product from the PATHS project.

	Paths products	Strengths	Weaknesses	Threats	Opportunities
Research results, Knowledge	PATHS as a paradigm for navigating cultural heritage collections	Rich potential to create exhibitions, to tell stories and berry pick content from deep collections; has a good fit with the ambitions of GLAM institutions to mediate their collections. Good potential as a tool to support research activities and to build document trails.	Currently limited to text and image based content. Depends on the quality of content that is actually available in CH collections. Are there software or system dependencies?	Many others are working on collection curation tools, from Google to other projects in the Europeana family.	Versions of the system to use in galleries which incorporate multimedia formats, where there is commercial investment in designing and developing such systems. Developing researcher tools for implementation by the research infrastructures
	Information retrieval	Technically sophisticated and robust information retrieval based on a hybrid between object oriented databases (MongoDB), an inverted index (Solr) and traditional SQL RDBMS (PostgreSQL). Linking based on global URIs rather than local serial number sequences, making the system Semantic Web-ready	The information retreival modules in PATHS are specific to the project and too closely tied to the Europeana ESE data model. This limits the application areas to content providers with ESE data at the outset.	The chosen approach is generic and can easily be adopted by third parties. Good to secure adoption of PATHS methodologies but more challenging in terms of gaining a competitive edge. Too many technologies involved.	Conduct further research to consolidate the information retrieval modules to use fewer software components and adapt them to work with any underlying data model where an information item can be anything identifiable by a URI rather than an ESE record.
	Recommender systems (EHU)	By analysing log-files the recommender system shows relevant content resources within the local information volume for any ESE 'item' in the PATHS subset of Europeana data	Limited log data available. No personalization, generic recommendations for any user.	Emerging functionality in the mainstream search providers including Google, Yahoo, Bing etc. may outcompete local solutions	Study the potential of building a custom logging mechanism that records all user behavior and indexes it to permit personalized recommendatins. Make the system independent of the ESE data model. Develop a component that may be integrated into any responsive user interface and offer it as a SaaS web service.
	Natural language processing	Innovative applications of Natural Language Processing technologies to analysis of cultural heritage collections. Development of novel task (typed similarity detection).	Limited in number of languages to which techniques have been applied (English and Spanish to date). Approaches that could be applied were restricted by the heterogeous nature of the Europeana collection.	Technology is rapidly moving and could be superseded quickly.	The technologies could be applied to other CH collections (or even other collections on similar topics). Techiques developed in project could be adapted for other langauges.
	Map-based' information visualization	Innovative way of visualizing and navigating hierarchical three structures through a virtual landscape where topic branches are visualied as islands where the size and placement is determined by the number of sub-topics and associated content resources.	Manual steps required to build the initial map tiles that shows the semantic landscape	No obvious threaths to this approach but the concept may be replicated fairly easily.	Continue the refinement of the methodology, consolidate the number of technologies required and contribute as visualization algorithm into existing open source data library.
Software products	Natural language processing tools for enrichment of metadata (EHU, USFD)				
	- Generation of links to wikipedia articles	Tools to automatically annotate CH documents with Wikipedia links.	Although the software is accurate, and performs well in comparison with alternative approaches, it is not 100% accurate.	Others could develop similar software or reimplement approaches developed in PATHS.	Could potentially be used by any CH provider to enrich their collections. Available within service being developed in LoCloud.
	- Generation of internal links	Tools to automatically identify links within collections.	Although the software is accurate, and performs well in comparison with alternative approaches, it is not 100% accurate.	Others could develop similar software or reimplement approaches developed in PATHS.	Could potentially be used by any CH provider to enrich their collections. Available within service being developed in LoCloud.
	Recommender system: real-time generation of links to relevant content in same collection (EHU)	Tools to provide suggestions based on past history of user.	Integration in existing search platforms requires investment in software development.	Widely used Information Retrieval software could offer this functionality.	Could potentially enhance the functionality of CH portals.
	Sentiment analysis for web content resources (I-SIEVE)	Tools to provde clients information on the present sentiment towards a topic, brand, product or concept based on NLP processing of web resources. Unique technology.	Relies on public APIs of search providers that are subject to change and versioning. Proprietary code-base and algorithms (but this may be and advantage for commercial exploitation)	Better suited for contemporary information where the 'buzz' is more active and the sentiment is descriptive of a situation that may be impacted by strategic changes and measures.	Develop the present bundling of services and automated analysis into a software as a service application. Build on the existing market potential and refine the methodologies and technology.
	Open Web API for application development (AVINET)	Rich functionality that is capable of powering sophisticated applications that offer more than simple search/browse interfaces Uses open standards for communication and data exchange. Based on market leading Open Source components.	Requires the comprehensive PATHS datamodel including the manually pre-processed Europeana data to work.	Requires a dedicated server if high user volume is envisaged	Modify the API so that it enables creation and discovery of PATHS on top of any underlying data model. This will increase the potential number of use cases greatly, whereby expanding the market potential. The market study suggests that 73% of the respondents find this technology interesting.
Services, metholologies	Content enrichment (USFD, EHU, AVINET)	Semi-automatic procedures that provides typed links between items in a collection and relevant background links to Wikipedia articles for ESE records.	Closely linked to the ESE data model	Alternative technology frameworks growing in number.	Make the process independent of ESE records as input and design a way of bringing data back into the source data model. There are many software toolkits in this domain but mostly with a high learning threshold and few serious service providers.
	User requirements specification (USFD, MDR)	Well-described methodology for user requirements analysis and functional user requirements specification that can be used as a basis for commercial consultancy services	Do not have a vessel to bring this type of services to market on a large scale.	There are many competitors in this domain and the envisaged clients, specifically within the cultural heritage domain, are likely to prefer local providers or providers with whom they have an existing relationship.	Disseminate and "open source" the methodology
	Web application user evaluation (USFD, MDR)	Well developed methodology for carrying out user evaluations in laboratory and field situations. Established expertise in collecting, analysing and reporting evaluation results	Requires a representative sample that cuts across the various user groups.	Requires access to the iLab for part of the methodology	Informal information suggests that not many people are currently carrying out user studies. There is a lot of interest in the results which suggests that there is potential to conduct further research.

Figure 24: Swot Analysis for PATHS knowledge, products and services

8 Exploitation Strategy

This chapter describes the high-level exploitation strategy by which the partners will ensure the long-term sustainability of PATHS results. The next chapter will elaborate the detailed exploitation plan for each of the identified PATHS products.

8.1 Mode of Exploitation

This section defines the different ways in which PATHS products may be exploited. We have chosen to divide the modes into three main strands: scientific, commercial and open source. Each strand includes a number of different measures that may be employed to secure that the products are adopted and brought forward.

While the prime objective of the project is to develop new knowledge that can form the basis for marketable products, it is also necessary to acknowledge that there are several steps between the outcomes of a scientific research project and a commercial product. Therefore, our modes of exploitation also include non-commercial strategies.

8.1.1 Scientific Exploitation

While some of the PATHS products have reached maturity where they may be brought to market, others require further research or have spawned potential for new independent research projects. Such products are suitable for scientific exploitation.

An obvious mode of scientific exploitation is to use PATHS research as the basis for future publications. The scientific partners of the PATHS consortiums, USFD and EHU assisted by MDR Partners have done extensive publishing and presentations of PATHS proceedings throughout the execution and at the time of completion, there are a number of publications in the pipeline.

The PATHS project operates within a wide array of technologies and several of the areas of research have the potential to be, or have already been, brought forward into new applied research and implementation projects. Several of the PATHS consortium partners are working jointly on new projects and through these, PATHS products that are not yet mature for the consumer market may be further refined towards that end.

8.1.2 Commercial Exploitation

Research projects often seek to demonstrate the feasibility of doing multiple complicated things simultaneously stretching the perceived boundaries for what is possible at the cost of performance and stability; many of the most successful business models on the Internet, on the other hand, show a tendency towards doing something very simple but doing it very, very well in a way that is robust, scalable and highly reliable.

For this reason, the PATHS platform as a totality is not marketable as it has too many dependencies on data, technologies and people to be possible to sell at a fixed price with a sufficient profit margin. Hence, our emphasis is on re-usable components.

PATHS products that are suitable for commercial exploitation must have a significant market potential, must have clear IPR, must be simple to define and easy to put a price tag on.

We have chosen to further subdivide commercial exploitation into three types:

- Sale of software licenses including all common software delivery paradigms: buy over-the-counter, download-and-install and software-as-a-service
- Sale of software library licenses for embedding of modular PATHS products within existing software products, whereby linking into existing business models and value-chains.
- Sale of consultancy services on hourly or project basis

8.1.3 Sustainability through Open Source

Since many of the PATHS products are built on Open Source licenses that propagate Open Source licensing requirements for upstream products, the bulk of the project results will be available as open source, free and open source or be entirely in the public domain.

Open Source is not the same as free and there are many actors in the market that live from sales of Open Source software (i.e. professional support licenses, customization, richer feature editions vs Community editions etc.) - but even free Open Source software provides a significant potential for Consultancy services whereby the PATHS products can enter into a software-life-cycle suitable for marketing. However, Open Source also gives two more opportunities for sustaining the project results:

- PATHS products may be made into new, independent Open Source projects on social development hubs like GitHub or SourceForge. By engaging a critical mass of developers and end-users, the products will be sustained without any costs incurring. Any PATHS partners or third parties will be free to develop commercial services based on the software.
- With the large amount of Open Source software projects in existence, a more likely way to sustain PATHS products is by contributing specific parts of the code to existing projects that already have a significant user base and/or critical mass of developers.

As an example to illustrate that this concept works in practice, not only in theory, PATHS consortium partner AVINET has built their commercial business model on bundling consultancy services with Open Source software since its start-up more than 10 years ago.

8.2 Responsibility and Implementation

It is the responsibility of the individual partners to implement the marketing strategies for PATHS products in accordance with the exploitation plan described in chapter 9. The consortium as a whole does not constitute a legal entity that may be used as a vessel for joint commercialization efforts, the establishment of such an entity will be assessed upon evaluating the results of the exploitation measures after one year.

In order to maximize potential use of the PATHS products all partners will when appropriate use PATHS knowledge, software products, services and methodologies in relevant Europeana contexts.

8.3 Costs and Revenue

Each partner is responsible for his/her own costs of bringing PATHS products forward. Any revenue resulting from the efforts shall first cover incurred expenses.

Any profits shall be divided between partners following a key defined for each product based on who has contributed to its realization.

9 Exploitation plan

This chapter describes the steps to be taken by the partners to ensure the sustainability of the PATHS products, both direct and indirect, beyond the project execution period.

- Research results, knowledge
    - PATHS as a paradigm for navigating cultural heritage collections (USFD, MDR) Working directly within the environment presented by Europeana and successor EU-funded programmes such as Horizon 2020, to disseminate, apply and further development the paradigm.
    - Information retrieval (USFD)
    - Recommender systems (UPV/EHU, USFD)
    - Natural language processing (UPV/EHU, USFD)
    - Information visualization (USFD)
- Software products
    - Natural language processing tools for enrichment of metadata (UPV/EHU, USFD)
    - “Topic map”, map visualization of thesaurus (USFD)
    - Recommender system: real-time generation of links to relevant content in same collection (UPV/EHU, USFD)
    - Sentiment analysis for web content resources (ISIEVE)
    - Open Web API for application development (AVINET)
- Services, methodologies
    - Content enrichment (USFD, UPV/EHU, AVINET)
    - User requirements specification (USFD, MDR)
    - Web application user evaluation (USFD, MDR)

9.1 Research results and knowledge

PATHS consortium members are actively seeking for research opportunities that will allow the continued research within scientific threads that have been created through the PATHS project.

The actions within scientific exploitation include publications, independent research and follow-up projects that adopt parts of the research potential developed through the PATHS project. The actual detailed measures have been described previously in the document including timelines and responsibilities; these are therefore not repeated in this section.

Several PATHS partners are jointly part of newly established projects where the content enrichment related technologies and web services developed through PATHS are further refined and developed in order that they in the future may mature into full-fledged commercial components.

9.2 Software products

In the exploitation plan, the major emphasis has been put on exploitation of software products as these are both tangible and easily measurable results from the project. The following sections describe the proposed exploitation strategies, actions and responsibilities

for the four highest ranked software products that the consortium members believe has significant potential.

9.2.1 Recommender system

9.2.1.1 Strategy

The recommender system knowledge and software specified in 4.1 and 4.2 can be exploited as knowhow, competence and a server (software as a service) and for commercial products and services.

UPV/EHU and USFD will continue to pursue recommender system research and applications as software services or consultancy.

9.2.1.2 Actions

- Publications to exploit knowledge produced in the project.
- Working directly within the environment presented by Europeana and successor EU-funded programmes such as Horizon 2020, to disseminate, apply and further develop the service.
- Check with consultancy companies (BicBerrilan, Bantec) market potential of recommender services, both as service (possibly embedded in another application) and as consultancy.

9.2.1.3 Responsibility

UPV/EHU (Eneko Agirre as local IP of the project, with the collaboration of the university OTRI - Ignacio Largo)

USFD (Paul Clough, as technical lead on recommendation systems, with input from USFD's Research and Innovation Services)

9.2.1.4 Timing

- Ongoing (Publications, H2020 program, Europeana)
- 2014 spring - meeting with consultancy companies.

9.2.2 Natural language processing tools for enrichment of metadata

9.2.2.1 Strategy

The content enrichment knowledge and software specified in 4.1 and 4.2 can be exploited as knowhow, competence and a server (software as a service) and for commercial products and services. This package would include all the following products:

- Identification of key entities within item descriptions
- Generation of internal links
- Generation of typed links
- Generation of links to Wikipedia articles
- Generation of identity links to Wikipedia articles
- Automatic generation of hierarchies

The LoCloud project is going to explore how to deploy a content enrichment service for Europeana. LoCloud is a Best Practice Network co-funded under the CIP ICT-PSP programme of the European Commission that will enrich the Europeana content by adding over 4 million digitised items from European cultural institutions.

UPV/EHU and USFD will continue to pursue content-enrichment research and applications as software services or consultancy.

9.2.2.2 Actions

- Publications to exploit knowledge produced in the project.
- Working directly within the environment presented by Europeana and successor EU-funded programmes such as Horizon 2020, to disseminate, apply and further develop the service.
- Deploy a content enrichment service in LoCloud (software as a service).
- Check with consultancy companies (BicBerrilan, Bantec) market potential of content enrichment services, both as service (possibly embedded in another application) and as consultancy.
- Check for potential applications with existing contacts within CH community and with Fusion IP (for USFD).

9.2.2.3 Responsibility

UPV/EHU (Eneko Agirre as local IP of the project, with the collaboration of the university OTRI - Ignacio Largo)
USFD (Mark Stevenson as local PI, with support from USFD's Research and Innovation Services, particularly Stephen Pyke)

9.2.2.4 Timing

- Ongoing (Publications, H2020 program, Europeana, LoCloud).
- 2014 spring - meeting with consultancy companies and discussion with Research and Innovation Services)

9.2.3 Open Web API for application development

9.2.3.1 Strategy

The Open Web API developed for the Paths prototype is knowhow, competence and a toolbox and for commercial products and services to be provided by AVINET in the future.

The following ongoing initiatives and projects will explore and utilise the results of the Open Web API process;

- Adaptive; Adaptive is a commercial-off-the-shelves web map solution developed and sold by Avinet. A completely new version of Adaptive is being developed and is due to be released in February 2014. A major asset of the new Adaptive is a comprehensive search mechanism, based on pre-indexed attributes and content from various sources. The search mechanism will be based on the services and components explored from PATHS.
- GeoStat; GeoStat is a completely new product Avinet is developing together with the county office of Rogaland and Sogn and Fjordane in Norway. The API will be re-used for search, exploitation of data and presentation.

- Linking the Norwegian registry for official place names with the Norwegian law and regulation database. A RDF based approach, built on the experience from the PATHS API will be utilised.

Avinet will continue to pursue innovative connecting services and is developing GIS as a service marked. Our strategy is to continue to share services, competence and APIs to provide first class products to our customers.

9.2.3.2 Actions

- Utilise the API for ongoing application developments.
- Internal Avinet training and utilisation of API components.
- Present API for partners for their exploitation.

9.2.3.3 Responsibility

The responsibility for the exploitation of the APIs within Avinet is internal (Head of Research and Development - Mr. Runar Bergheim).

9.2.3.4 Timing

- Release of Adaptive 3.0 ultimo January 2013

9.2.4 Sentiment analysis tools for web content

9.2.4.1 Strategy

Sentiment analysis in the cultural heritage sector is expected to be used as a crowdsourcing recommendation system. The plethora of artefacts in digital representation is expected to rise exponentially in the immediate future. The nature of the Web, with all digital objects being born equal, as well as the “web culture” rapidly adopted by younger generations plays down the authority of publisher when compared to public opinion.

Of course, adoption of such viewpoints from memory institutions is not yet abundant and the task of educating this market segment is not an easy one.

The case study of PATHS is a good starting point and i-sieve will put all effort required to have the system up and running for long after the end of the project so as to act as a showcase of possibilities. This will also give i-sieve the opportunity to trim its systems in such a way that they will be optimal for the cultural heritage sector.

In parallel, i-sieve have planned a set of free sentiment analysis reports on the content of stakeholders, so as to entice them into looking at the standing opinion for their own content. It will begin this marketing effort with the memory institutions that helped with the evaluation of the system, in the sense that these are the potential clients with a less steep learning curve for the utility of such systems. Armed with their feedback, i-sieve will then approach other major stakeholders with the same offer.

9.2.4.2 Actions

- Reserve and maintain resources for the continuity of service for the PATHS prototype
- Perform and present our free reports to up to 4 of the memory institutions involved in the evaluation stage
- Approach other major stakeholders

9.2.4.3 Responsibility

Konstantinos Chandrinos, Managing Director of i-sieve technologies

9.2.4.4 Timing

Resources to maintain prototype – ongoing

Free reports for 4 institutions – concluded by March 2014, presented by April 2014

Approaching of other key players – to begin in the autumn of 2014

9.3 Services and methodologies

In addition to software products, PATHS has also developed a number of methodologies where expert tools are bundled with consultancy service. This is applicable in cases where the processes are very complex and/or where available tools can only be utilized manually by expert users.

In addition, the PATHS project has also further developed and refined well-known methodologies for user-requirements-gathering and end-user testing. This section describes how these methodologies and services will be brought to market.

9.3.1 Strategy

The exploitation strategy is a combination of commercial exploitation and Open Sourcing well-described methodologies. This will be done through dissemination of the methodology reports presented on the project web site, and further disseminated through PATHS partners' networks.

9.3.2 Actions

- Write popular article 'on top of' formal deliverable in order to provide an easily accessible intro to the methodologies
- Probe and develop service concept for expert-assisted content enrichment consultancy services for the cultural heritage sector and beyond
- Adopt methodologies into day-to-day working procedures among partners where applicable.

9.3.3 Responsibility

The responsibility for exploiting PATHS methodologies to develop consultancy services lies with each partner but with specific responsibilities assigned to:

- All partners for making the deliverables accessible through search engine optimization linking and development of intro articles.
- EHU for the follow-up of consultancy services within content enrichment.
- SME partners AVINET, i-Sieve and MDR Partners for adopting methodologies into day-to-day operating procedures of commercial entities.

9.3.4 Timing

The actions outlined above will take place in Q1/2014

10 References

L. Padró, M. Collado, S. Reese, M. Lloberes and I. Castellón. 2010. FreeLing 2.1: Five Years of Open-Source Language Processing Tools. Proceedings of 7th Language Resources and Evaluation Conference (LREC 2010), Valletta, Malta. May, 2010.

N. Aletras, 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.

E. Agirre, 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

E. Agirre, 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

E. Agirre, N. Aletras, A. Gonzalez-Agirre, G. Rigau and M. Stevenson. 2013. UBCUOS-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

E. Agirre, A. Barrena, K. Fernandez, E. Miranda, A. Otegi and A. Soroa. 2013b. PATHSenrich: A Web Service Prototype for Automatic Cultural Heritage Item Enrichment. In TPDL 2013, Lecture Notes in Computer Science volume 8092, pages 462--465.D7.4:

D. Milne and I. Witten. 2008. Learning to Link with Wikipedia. In Proc. of CIKM 2008, Napa Valley, California.

S. Ponzetto and M. Strube. 2011. Taxonomy induction based on a collaboratively built knowledge repository. Artificial Intelligence, 175 (9-10), 1737-1756.

S. Fernando 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.

M. Hall and P. Clough 2013. Exploring Large Digital Library Collections using a Map-based Visualisation, In Proceedings of The International Conference on Theory and Practice of Digital Libraries (TPDL 2013), pp. 216-227.

P. Clough, A. Otegi, E. Agirre and Mark Hall. 2013. Implementing Recommendations in the PATHS System, Proceedings of the Second International Workshop on Supporting Users' Exploration of Digital Libraries.

D5.1 Evaluation of the first PATHS Prototype (2012). PATHS deliverable.

D5.2 Evaluation of the second PATHS Prototype (2013). PATHS deliverable.

D5.3 Field Trials of the PATHS Prototype (2014). PATHS deliverable.

M. Olensky, 2012. Market study on technical options for semantic feature extraction. Europeana V2, Deliverable 7.4