ECP·2005·CULT·038206

Discovery

D1.8 Final Report

1 November 2006 – 31 October 2009

Deliverable number	*D1.7*
Dissemination level	*Public*
Delivery date	*31 October 2009*
Status	*Final*
Author	*Paolo D'Iorio (CNRS)*

e*Content*plus

Table of Contents

1 Project Objectives

The Discovery project has two main components:

Philosource, a federation of semantic digital libraries in the field of philosophy.

Philospace, a network of personal desktop applications used to enrich the content of Philosource with semantic annotations.

The *Philosource* federation contains an extensive, multilingual collection of reliable scholarly editions of philosophical texts, high quality reproductions of primary sources and a rich archive of videos including lectures and interviews featuring leading contemporary philosophers. Each of the Web sites of the Philosource federation is dedicated to a particular body of work and all materials published in Philosource have stable and persistent Web addresses to ensure the reliability of scholarly references. All of the content in Philosource is freely available via any common Web browser and is fully interoperable with Europeana.

Philosource is thus like a union of independent countries, brought together by their commitment to common values (free access, collaboration, consistency of sources, standards of excellence, peer review), innovative Web technologies, and the constant exchange of ideas. Think of this material as a vast territory ready to be explored, described, and mapped out. The original configuration remains intact, even as the territory expands with each new contribution.

You can travel through the territory at will but you can also design new maps and author guidebooks for others to read and respond to. In fact, on top of this network of Web sites lies Philospace, a desktop application that runs on a scholars' personal computer and provides a collaborative environment in which to browse, study, and enrich the content published in the Philosource federation. Thanks to the use of Semantic Web technologies it allows users to make ontology-based contributions such as comments and annotations and to share them with other users or merge them with other external Semantic Web sources. Philosophy has often been thought of as a journey or exploration. Discovery aims to find out what this voyage could look like in the 21st century.

discovery
digital semantic corpora for virtual research in philosoph

2   Consortium

2.1   Institut des Textes et Manuscrits Modernes, Paris, France (CNRS)

Role: Coordinator, content provider (Friedrich Nietzsche). The Institut des Textes et Manuscrits Modernes (ITEM) is a joint research unit of the Centre National de la Recherche Scientifique (CNRS) and the Ecole Normale Supérieure (ENS), Paris. ITEM is the main centre (in Europe and throughout the world) for the study of genetic criticism. ITEM has been engaged for years in developing methods to study philosophical and literary texts genetically, i.e., examining the evolution of texts and ideas from a first mentioning in a notebook, through different stages of rewriting, to the published version. Twelve researchers and fifteen research and administrative assistants work full time on genetic projects, with the help of seventy two academic associates from different universities and eleven associated curators from the Bibliothèque nationale de France (BnF) on authors like Flaubert, Zola, Nietzsche, Proust, Valéry, Joyce, Celan, Sartre, and twentieth-century Latin-American literature. Recently ITEM has opened researches lines dedicated to French literature from the West Indies and Africa, and to visual art and cinema. Its documentation centre houses the most comprehensive and up-to-date collection of critical works devoted to genetic criticism as well as many facsimiles (photographic or digital) from manuscript archives throughout the world (www.item.ens.fr). During the third year, the project moved to the Maison Française d'Oxford (www.mfo.ac.uk), which opened a new research area dedicated to Digital Humanities.

2.2   Wittgenstein Archives at the University of Bergen, Norway (WAB)

Role: Content provider (Ludwig Wittgenstein). The Wittgenstein Archives at the University of Bergen (WAB) is a research infrastructure in philosophy, editorial philology and text encoding. It is probably best known for the publication of "Wittgenstein's Nachlass: The Bergen Electronic Edition" (OUP 2000). This edition (on 6 CDs) contains all of the manuscripts of Wittgenstein's *Nachlass* in facsimile and transcription. WAB is embedded in the Research Group for Text Technology at the Department of Culture, Language and Information Technology (AKSIS) at Unifob AS, owned by the University of Bergen. AKSIS' main focus includes language and text technology, ICT and learning, and culture research. The Text Technology Group develops information technology solutions for text encoding, digital editorial philology, and classification and organisation of massive text collections. They focus on integrating the possibilities and challenges of the knowledge society into their projects, with emphasis on e-learning, knowledge bases, generalized access to technology, and knowledge dissemination. The group is part of a large, collaborative network of international partners (wab.aksis.uib.no).

2.3   Net7 Internet Open Solutions, Pisa, Italy (Net7)

*Role: Technology provide*r. Net7 is an SME (Small and Medium Enterprise) based in Pisa, Italy, whose core business is the development of Open Source and Free Software applications supporting scientific projects. Its research group dedicated to Humanities, Digital Libraries and Cultural Heritage collaborates with major international research institutions on several large-scale international research projects, including an EU COST Action (www.cost-a32.eu) and a *Marie Curie Industry-Academia Knowledge Transfer* project. In collaboration with the University of Pisa's Computer Science Department, Net7 is also supporting a Master's

program in Open Source Management. Projects include *Opera Liber*, a Web application built with Semantic Web technologies such as RDF and OWL and used to support the study of Italian opera libretti and *HyperJournal*, an Open Source Web application for facilitating the Open Access publishing of electronic journals on the Web. Net7 is also active in the fields of functional languages, Content Management Systems, and E-Government (www.netseven.it).

2.4 Lessico Intellettuale Europeo e Storia delle Idee, Rome, Italy (ILIESI)

Role: Content provider (ancient Greek philosophy, philosophical and scientific literature from 16th to 18th Century). The *Lessico Intellettuale Europeo e Storia delle Idee* is a CNR institute whose mission is to conduct research in the field of European philosophical and scientific thought or, more generally, in the field of the history of ideas with a focus on Antiquity and Late-Antiquity, and Early Modern Age (from the Renaissance to the Enlightenment). Special attention is paid to the evolution of languages and to philosophical lexicon. ILIESI has extensive experience in the automatic treatment of texts, particularly in the fields of computational philology and lexicography. In more than 40 years of activity, the ILIESI has set up a very large database of philosophical and scientific texts in several languages and published over one hundred volumes (www.iliesi.cnr.it).

2.5 Università Politecnica delle Marche, Ancona, Italy (UNIVPM)

Role: Technology provider. The 3-Media Labs at Università Politecnica delle Marche pursues a diverse research agenda including signal processing, neural networks, telecommunications networks, knowledge representation, Web technologies, and general computer science. SeMedia, one of the two working groups within 3-Media Labs, has focused increasingly on Semantic Web research, and in particular on integrating multimedia into Semantic Web structures. Current SeMedia projects include DBin, an open-source Semantic Web knowledge management application (http://dbin.org) and Semantic Web Pipes, a Web based data aggregation and transformation tool based on RDF (http://pipes.deri.org).

The SeMedia research group has developed innovative contributions to the Semantic Web P2P field with the RDFGrowth algorithm for replication of "aspects" of knowledge across peers and methodologies for applying digital signatures on RDF graphs. SeMedia has also experimented with configurable Graphic User Interfaces for domain specific Semantic Web applications.

2.6 RaiNet, Rome, Italy (RAI)

Role: Content provider (audiovisual lectures and interviews). RAI, Radiotelevisione Italiana is Italy's public broadcast network, and operates three terrestrial television channels and three radio channels, as well as satellite and digital services. RaiNet makes RAI content available on the Web through the RAI.IT portal and develops original content for the Web. The portal also provides public information about institutions and government, including streaming audio and video from the Parliament. RAI has a large archive of sound and video files featuring prominent philosophers discussing a wide range of philosophical topics. Key contributions to philosophical research include the videos from Rai Archive (*Multimedia Encyclopaedia of the Philosophical Science* and numerous TV programs related to this important collection of 1,500 interviews and lessons) as well as a series of new interviews by RaiNet Editorial Staff (www.rai.it).

3 Project Results

3.1 Philosource: open access to scholarly content in the field of philosophy

The Philosource federation is an extensive and coherent collection of open access scholarly content from classical to contemporary philosophy. While there have been many projects that focus on making archival or scholarly content available, the primary task of the Philosource federation is to make a comprehensive corpus of quality materials *usable* in easy, efficient, and scholarly ways. Through a deep dedication to both comprehensiveness and usability, Philosource aims to make working with these materials online even more rewarding and beneficial than working with them offline.

The Philosource federation provides access to a wide range of key texts and materials in Western philosophy, with a special emphasis on Greek and Latin philosophy, early modern philosophy, Nietzsche, and Wittgenstein. It includes works in Greek, Latin, French, Italian, English, and German and from different periods of these languages. The material is rich and diverse in geography, historical period, language, and philosophical adherence. It contains elaborate works presenting final substrata, and also includes—for Nietzsche and Wittgenstein—their very first notes and raw drafts. Both previously published and widely used works and editions are represented, as are new editions and materials for which access up to now has been difficult or fragmented. The philosophical genres range from treatises in logic and philosophy of science to explorations into the functions and value of literature and the humanities. The text genres employed include dry argument, logical and mathematical notation, and strict treatises, but also lectures (both academic and popular), dialogues, graphics and drawings, poetic works and essays of a loose form. A wide range of different working and revision paths are provided, including notebook writing, fair copy, cutting and decomposition, collage and rearrangement, and paths from first drafts and sketches through selection, revision and rewriting to elaborate final distillations. Editions included represent both readers' and scholarly critical editions, among them genetic editions. Audiovisual materials from RaiNet add significant value to the resource with expert interviews with some of the most renowned philosophers of the 20th century.

The Philosource federation provides access to more than 65,000 units of philosophical content. ILIESI has made the entire Presocratics, Socrates and Digenes Laertius corpora available. From early modern philosophy, they have included the key texts of rationalism and early Italian Humanism. CNRS will make available a huge part of Nietzsche's philosophical archive (published works, proofs, drafts, preparatory manuscripts) as well as the reference and critical edition of works and unpublished notes edited by Colli/Montinari. From Wittgenstein, 5,000 pages of facsimile images and related transcriptions and published editions from four different periods are now represented, including his *Notes on Logic* (1913, English) and *Lecture on Ethics* (1929, English), notebooks from the early nineteen thirties (German), as well as the famous *Big Typescript* (1933, German)—Wittegenstein's *Summa philosophiae*—with its predecessor TS 212, and the *Brown Book* (English), one of the most quoted and utilized pieces of Wittgensteinian philosophy. RaiNet have provided 300 video lectures and interviews featuring leading contemporary philosophers and related to the texts provided by the other partners. These pieces were selected to be published first not only because they represent a broad range of user needs and interests but also because they are some of the finest and most "wanted" of Western philosophy.

3.2 Ontologies

The Philosource federation use a common ontology, called the *Scholarship Ontology* to express the distinction between research objects (primary sources), research results (secondary sources), and authors of both (people). The *Scholarship Ontology* also expresses the general relationships between these sources and their authors, such as "related to", "describe", "criticize", "comment", etc. It has been specifically designed for simplicity and model compatibility with other standards and specifications, such as the Functional Requirements for Bibliographic Records (FRBR) and the Dublin Core metadata standard.

Along with the Scholarship Ontology, each node of the Federation use narrower *domain source ontologies* to express the concepts used by the concerned authors and their relationships. Formalized in the RDF Schema and OWL, these philosophical ontologies facilitate the semantic enrichment of the content providing the users with the terminology appropriate to the content of the Website they are reading. To this end, each Philosource Website exposes to Philospace and to other semantic Web applications a set of public recommended ontologies to be used in semantic annotations. Scholars can use the suggested ontologies or, thanks to the flexibility of the Philospace software, are able to extend them with new concepts and relationships or to design their own ontologies thus enabling personal annotation environments.

3.3 Semantic Enrichment

Semantic enrichment is a broad concept, which refers to the integration of metadata or commentaries in digital content, particularly in scientific and scholarly contexts. Discovery has demonstrated the possibility of several different types of semantic enrichment applied to philosophical texts and realized significant concrete examples of enrichment on chosen subsets of content.

1) The *collaborative annotation of content* is a first and simpler form of semantic enrichment. The scenario here is the existence of a philosophical community, which intends to work on a certain text in a collaborative way. Using the combination of reliable scholarly texts contained in the Philosource federation on the one hand, and the flexibility of the Philosource software on the other, scholars will be able to individually annotate the content, to share their annotations with colleagues, and to structure their commentaries using the preloaded philosophical ontologies available for each of the sites of the federation. CNRS has produced an example of collaborative annotation related to Nietzsche's work in order to encourage external scholarly communities to follow and adapt this model to their needs.

2) *Knowledge Representation*. This more rigorous form of semantic enrichment attempts to represent in a machine-readable way the philosophical concepts contained in the texts. To this end, it uses specialized ontologies and explored what kinds of interrelation could be established and which forms or reasoning would be allowed. The project provided several different examples of knowledge representation:

2a) *One concept in the complete works of one author*. CNRS prepared a domain ontology dedicated to the concept of nihilism and enriched all Nietzsche published and unpublished works applying this ontology to the relevant passages. This will permit users to navigate Nietzsche texts following the structure of the concept of nihilism with his entire ramification.

2b) *All concepts in a subset of works of one author*. All important concepts contained in the texts published in Wittgenstein Source (ca. 5,000 pages of Wittgenstein Nachlass) were enriched using a specific domain ontology. Because the Wittgenstein Archive initially used the *Protegé* software to perform this type of semantic enrichment, this experiment also permitted the testing of an additional important feature of Philospace, that is the possibility to import pre-existing knowledge bases based on RDF. This feature could be of interest for users who have already started a form of semantic enrichment with other software and would like to switch to Philosource without loosing the first results of their work.

2c) *Several concepts in different authors*. The ILIESI Websites publish works belonging to different authors coming from different ages of the history of philosophy. This provided the occasion to experiment with a different type of knowledge representation, that is the use of an upper, multilingual ontology to map concepts belonging to different works of different authors written in different languages. This form of enrichment will allow users to navigate a multilingual collection of philosophical texts following a semantically structured pathway. From an interpretative point of view, this provides an interesting opportunity to see how different authors of different ages and languages have used a certain concept.

3) *Interrelated multimedia contextualisation*. Philospace, finally, also permits the creation of structured links between the transcriptions of lectures and interviews hosted by Sophiavision to all related texts or concepts spread in all other Websites of the Philosource federation.

3.4   Software development: Talia

The Philosource federation is powered by Talia*,* an innovative new software platform specifically designed for the needs of scholarly research in the humanities. Talia combines all of the functionality of a digital library with an electronic peer-reviewed publisher, creating a powerful online workspace for humanities scholars. Talia is able to handle a wide range of resources including texts, images, and videos. All of the resources published are identified by a stable URI, meaning that documents can never be removed once they are published and are maintained in a fixed state in perpetuity. Discovery's commitment to persistent URIs and long-term preservation gives scholars the assurance that their citations of materials in the Philosource federation will persist, thereby removing one of the greatest barriers to adoption of digital resources by humanists.

Beyond its innovation in meeting the needs of scholars, Talia is pioneering the adoption of Semantic Web technologies—including RDF and the strategic use of ontologies—to create a state-of-the-art research and publishing environment. In this first instance of *Talia*, users can access facsimiles, critical editions, and video interviews by browsing scholar-created ontologies.

Talia is built on the *Ruby On Rails* (www.rubyonrails.org) Web development framework, a popular open-source framework known for its agility, speed, and sustainability. To enable powerful and advanced searchers into the transcriptions of texts, each Talia instance has been integrated with a remote XML search engine based on *eXist*. For working with high-resolution facsimile images, Talia integrates the *IIPImage* technology, allowing users to zoom into images up to a high level of detail without forcing them to download large files.[1] An

---

[1] See for example http://merovingio.c2rmf.cnrs.fr/iipimage/IIPHTMLDemo-Globe.html, which shows a 10,7 GB image of the earth that is served directly to the browser.

instance of Talia has been deployed on each content partner's server and populated with their content. (www.talia.discovery-project.eu).

3.5   Software customization: Philospace

Philospace is a desktop application that allows users to annotate Web pages with semantically structured information. Philospace is mainly targeted to scholars and, more specifically, it has been designed to address the Philosource scenario. However the ideas and functionalities can be used in a variety of other contexts.

Through an internal Web browser, users can access the content published in the Philosource federation—or in any other Web page—simply by typing in the correct URL. Philospace then allows the user to select any part of a Web page, e.g. a sentence, an image, or even a single word, and add a semantic annotation simply by choosing the 'Add Note' option. This will take the user to a graphical wizard where all of the relationships, classes and instances defined in the pre-loaded ontologies can be used to express a semantic statement. For example, a sentence can be marked at defining the concept of 'Freedom'. This set of connections actually form what we call the *knowledge graph*, in which the annotated resources are the nodes and the relations among them are the edges.

In Philospace, each user has a *personal notebook,* which stores the comments and semantic connections that the user has created among Web pages and entities (persons, concepts, places) of interest. The annotations created with Philospace are by default, kept private, stored in the user's local file system, and displayed contextually when the Web page is visualized in the Philospace browser. A user can also decide to make his semantic notes public through the 'Publish' option on the preferences page. *Public annotations* are stored on the Philospace Web site and are retrievable by a stable URL. Once public, users can import all of the annotations made by other members of the Philospace community, making them visible when the page is displayed in the browser. When a user imports a notebook from the Web his knowledge graph grows, showing more information and connections and Philosource becomes a useful tool to discover new knowledge while browsing Web documents.

Different kind of user communities have different needs in terms of annotation capability, in particular regarding the terminology and vocabulary of concepts depending on the specific domain of interest. In Philospace this is addressed by allowing users to import *multiple Semantic Web ontologies*. Furthermore, in order to implement a pragmatic solution to reach ontologies consensus among users, Philospace allows Web sites themselves to automatically recommend ontologies to their visitors.

Philospace is completely open source and available for download, along with an introductory user guide, at: www.dbin.org/swickynotes/philospace. It is based on a more generic platform for semantic Web annotation that is called SWickyNotes. Currently supported platforms are Windows and Mac OS X.

3.6   IPR Issues

Discovery is a fully Open Source and Open Access Project. It guarantees public access to Websites, software, and content produced or published within the Discovery federation at no charge. No technological means will be used to limit or control access to or copying of software or content.

All software produced during the project has been released under the terms of the MIT license, as certified and published by the Open Source Initiative Web Site (http://www.opensource.org/licenses/mit-license.php). The MIT license allows anyone to incorporate code and results into other commercial or non-commercial projects.

At the content level the project will use a legal framework specifically intended to promote the principles of Open Source and Copyleft for the publication on the Web of primary and secondary sources in the humanities. The license used, the Creative Commons Public License Attribution Non-Commercial Share-Alike (CCPL-BY-NC-SA, http://creativecommons.org/licenses/by-nc-sa/3.0/legalcode), grants access and reproduction rights to everyone, while at the same time protecting the integrity and the paternity of the works. All the grants arising from the license refer to non-commercial use, thereby limiting the commercial exploitation of the enriched content and reserving such rights for the author.

4 Targeting Users & their Needs

4.1 General Strategy

In order to gain a more immediate understanding of the current working styles and habits of scholars in philosophy and the broader humanities, a team at the Oxford Internet Institute (OII) has begun engaging in an in-depth user needs assessment strategy. The principal aim of these activities is to gain insights into the current online working habits and prospective future needs of a population identified as potential Discovery project users. The data gathered during the assessment has been used to influence the development of the Philospace and Philosource software.

There are several components to the assessment plan. The first was an initial focus group, followed by a large-scale non-probabilistic, online survey, conducted in the Spring of 2008. Data from these activities were used to generate a series of wireframes and mockups used to guide subsequent phases of assessment. These mockups then formed the core of a series of focus groups conducted both virtually and in person (in several countries) aimed at a close analysis of the usability of the Discovery Project interfaces. Finally, a questionnaire dispersed after the release of Nietzsche Source at the *17th International Conference of the Friedrich Nietzsche Society of Great Britain and Ireland* by the Faculty of Philosophy and the University of Oxford in September 2009, allowed for a unique opportunity to gain valuable feedback from a select group of current and potential users.

4.2 First Focus Group

All of this work builds on a first focus group, which took place at the philosophy department of the university of Bologna. This initial enquiry was used to investigate four basic questions about the use of the Internet for scholarly research. 1. What is the role of the Internet in the every day working and research environment of scholars and students? 2. How is the Web (and Web 2.0) impacting communications among colleagues and the genesis of ideas? 3. How is the Web changing the practice of scholarly publishing? 4. What is the role of preprints and other forms of circulation of unfinished works among colleagues and research teams in changing collaborative practice?

This first focus group provided a clear sense of how young scholars in the human sciences approach the Web in general: while they can hardly avoid making heavy use of it and are eager to access more primary and secondary sources, they do not really want to recognize the overall importance of the Internet in their work. The biggest obstacles to uptake seem to be the perception of trust in information found on the Web, and, for those already expert in their field, finding information that can satisfy their precise needs. The results of this focus group were quite unexpected. Instead of finding widespread acceptance of new technologies by young scholars, we found them quite tied to 'traditional' practices of scholarly publishing and not very open to experimenting with the possibilities offered by the medium. Our interactions with the group confirmed the existence of what Jean Claude Guedon called a "Doctor Jekyll and Mr. Hide syndrome," i.e., that there is a fundamental difference between the behaviour of scholars as readers and scholars as authors. The participants of the focus group confirmed that they make use of a wide range of online sources in their research, like Wikipedia, but that they would never use these to publish their findings. In particular they seemed especially concerned with issues of trust, reliability and above all protection of their intellectual property.

4.3 Large Scale Project Survey

The Discovery Project merges online and offline environments in unprecedented ways, combining the traditional activities of scholarship with a dynamic and open online environment now dominated by non-scholarly social networks. In order to gauge acceptance and guide development of the Discovery online environment, a large-scale domain based survey was conducted to assess (1) current use of Information and Communication Technologies (ICTs) for scholarship-based activities (i.e., activities related to the work of the scholars but not including personal use of ICTs) by the target audience, (2) current perceptions about Web communities (scholarly and non-scholarly), (3) potential barriers to the adoption of online scholarly communities and work environments.

The survey contained 58 questions and was designed to take no more than ten minutes to complete. The questions were divided into four categories:

1. General Working Style and Habits
2. Participation in Online Activities Related to Your Work
3. The Future: What if?
4. Background and Demographics

The survey was made available online from 24 March to 13 April 2008 and attracted 512 responses. The target audience was faculty, researchers, and students primarily in philosophy and but also in the broader humanities. Potential survey respondents were identified individually as well as through organizations and listservs and were all identified as a) potentially using or being interested in the manuscripts and archives of philosophers or b) using or being interested in manuscripts and archives in general.

Executive Summary of Findings

About the Respondents

94% of the survey respondents are from the humanities; most (51%) are in philosophy. 86% are engaged in teaching or research activities at an academic institution. An online survey generally does not reach those less inclined to use ICTs in their work, and not surprisingly, the respondents here are for the most part quite technically savvy. Close to half of respondents maintain a Website or blog and most of those created the site themselves.

Current Working Style and Habits

Respondents work largely in a hybrid environment. There is still a significant reliance on print materials in all categories of resources used. Most respondents access their research materials both online and offline. Questions regarding preferences for "organizing one's thoughts" also indicate a hybrid environment. While most respondents use a computer to write their articles and papers, activities such as taking notes are not always facilitated by ICTs. Several respondents mentioned the citation management tool Zotero as one of their favourite online tools.

Searching and Finding Resources

Reference tools and searching for secondary sources are the most common online activities amongst the respondents. Respondents use online library catalogues more often than Google Scholar, but general Internet search engines (Google, Yahoo, etc.) are still the most heavily used search tools. Current use of online primary archival resources is only slightly higher than current use of CD-ROMs, but interest in seeing more of these materials online is high. Many respondents cited online subscription databases such as JSTOR, Early English Books Online, and various online reference sources as making their work significantly easier.

Collaboration

Collaboration amongst respondents is higher than one might think, given the stereotype of the 'lone humanities scholar', but a majority never use online collaboration tools. The need for more collaboration in the humanities—and collaboratively building humanities-based tools— was cited in the comments of several respondents.

Publishing Online

Most respondents do not currently make their work—published or unpublished—available online. The reasons for not publishing online are widespread, but copyright restrictions by publishers, a lack of respect for online publications, and not having enough time are the most common. Many said they do not make their work available online because their publisher already does. This assumption is troubling, though, as most publishers only make articles available online through subscription databases.

Participation in Online Communities and Annotation Activities

Current online activities are generally passive. Respondents search for materials online, but generally do not post comments to blogs or participate in online collaborative discussions. The number of people who say they would go online for a community in their field is high, but incentives for doing so favour finding primary resources over collaborative or participatory activities. Comments about online annotation activities indicate a strong desire for the ability to annotate online sources, but a hesitation to share annotations and an even greater hesitation to comment on other's annotations. Respondents generally favour a high level of granularity in access control to their annotations—i.e., being able to share some annotations and not others.

Peer Review

Peer review is extremely important to respondents; informally reviewed content is not of importance. This was echoed in several respondents comments—citing a need and desire for material that they can trust, without having to sort through a generic rating system.

Other

In principle, respondents support open source models, and dislike having to register for a site or use a password; that being said, almost all of the online resources mentioned by name as

being heavily used or useful are subscription databases (JSTOR, Early English Books Online). Several respondents expressed a desire for more available audio files or podcasts of lectures.

4.4   Subsequent Focus Groups

In order to conduct focus groups and to gather responses and reactions to the organization, design, and layout of the Philosource federation nodes, a series of detailed mock-ups were created allowing users to simulate a path through the different features. These images were used to guide a series of focus groups with key user groups.

Blog-style Focus Groups

In order to reach Nietzsche scholars around the world, a blog interface was used to lead a discussion about the interface. Initially 25 participants from around the world were hand-selected and invited to participate. The majority of these participants were faculty, but some advanced graduate students were included. Despite the personalized invitations, only a few participants responded and the focus group was re-run to a more open audience. While this brought in more participants, it proved a difficult environment to spark a discussion. In order to further address the need for feedback from potential users, a select group of participants associated with the Wittgenstein Archives at the University of Bergen (WAB) were asked to participate in the focus group. Of all of the methods employed in gathering feedback from users, this provided the most in-depth responses. While the participants were not able to respond or react to each other's comments, they were able to engage with the moderator in a familiar, non-threatening environment (email) that had low technical overhead. They were able to browse the site and complete the questions at their own convenience.

In-Person Focus Groups

In contrast to the email responses, a series of in-person focus groups generated immediate responses that were spontaneous and participatory. Keeping the group size between five and ten allowed everyone to have a chance to respond while still allowing for enough diversity so as to foster a dynamic discussion. In Ipswich, Massachusetts (just outside of Boston), a group of electronic publishers and librarians were assembled at the headquarters of Ebsco Publishing. In early January of 2009, a focus group was held during a conference in Cambridge, UK focused on, "Building a Humanities Collaboratory".[1] This diverse group consisted of practitioners in the Digital Humanities, and provided an opportunity for feedback from colleagues engaged in building digital collections and online research environments.

Nietzsche Source Questionnaire

After the release of Nietzsche Source, the coincidence of the hosting of *17th International Conference of the Friedrich Nietzsche Society of Great Britain and Ireland* by the Faculty of Philosophy and the University of Oxford, allowed for a unique opportunity to gain valuable feedback from a select group of current and potential users. Scholars from over 25 countries were in attendance in addition to those who are undoubtedly the most well known in this field. With support from the Oxford E-Social Science Project[2], a questionnaire was drafted and handed out to a sample of conference attendees. The main results of the questionnaire can

---

[1] See: www.ebscohost.com/thisTopic.php?topicID=433 and www.crassh.cam.ac.uk/events/751.

[2] The OeSS is housed at the Oxford Internet Institute, see: www.oii.ox.ac.uk/microsites/oess.

be summarized as follow: A majority of respondents had already heard of Nietzsche Source and two were already regular users. All respondents agreed that Nietzsche Source will be important to their field, particularly to their research. Only 50% of respondents agreeing about its importance for teaching. Users would overwhelmingly (100%) recommend the site to a colleague.

Conclusions

It is evident from all of the user needs activities conducted by Discovery that the time is right for the release of large collections of primary resources in philosophy. Scholars are eager for the availability of more primary resources in their field. Response from all user needs participants was unanimously positive regarding the need for these materials.

Reactions to the availability of tools to work with primary sources that will be made available through the Scholar Mode of Philosource and through Philospace was also positive, but users show some initial hesitance about what can actually be done with such tools. The hesitance implies that it is important to provide a familiar environment first, before introducing new elements to users.

5 Underlying Content

While Philosource is designed to be a continually expanding resource, it begins with a critical mass of material. Here is a brief overview of the content of Philosource:

Collection	Textual, Image or Video Units
Ancient Greek Philosophy	8,902
Philosophical and Scientific Texts	5,766
Nietzsche	39,000
Wittgenstein	12,172
Audiovisual Lectures and Interviews	308
Total	**65,840**

5.1 Ancient Greece

Three monumental bodies of work form the core of the ancient material in Philosource:

- a complete electronic edition of the fragments and testimonia of the *Presocratic philosophers*, based on *Die Fragmente der Vorsokratiker* edited by Diels and Kranz, in ancient Greek with translations into German and Italian (presocratics.daphnet.org);
- a complete electronic edition of all testimonia related to *Socrates and the so-called Minor Socratics*, based on Giannantoni's *Socratis et Socraticorum Reliquiae* and including, in addition, the text of Aristophanes' *Clouds* and Xenophon's Socratic writings, all in ancient Greek (socratics.daphnet.org);
- the complete text of *Diogenes Laertius' Lives of the Philosophers*, in ancient Greek with accompanying Italian translation (laertius.daphnet.org).

This material has been edited and published by the Istituto per il Lessico intellettuale europeo e Storia delle idee (ILIESI), CNR, Rome.

5.2 Early Modern Philosophy and Science

ILIESI has also published a selection of modern philosophical and scientific texts in Latin, Italian, and French, including 52 works by 7 major authors from the 16th to the 18th centuries (modern.daphnet.org):

- G. Bruno, *De l'infinito, universo et mondi, Spaccio de la bestia trionfante* and others
- R. Descartes, *Meditationes de prima philosophia*, *Passions de l'ame* and others
- B. Spinoza, *Tractatus politicus*, *Ethica ordine geometrico demonstrata* and others
- G. W. Leibniz, *De primae philosophiae principia* (Monadologia), *Principes de la nature et de la grâce fondés en raison* and others
- G. B. Vico, *Principj di una scienza nuova, De uno universi iuris principio et fine uno* and others
- A. G. Baumgarten, *Meditationes philosophicae de nonnullis ad poema pertinentibus*
- I. Kant, *De mundi sensibilis atque intelligibilis forma et principiis dissertatio* and others

5.3 Friedrich Nietzsche

Nietzsche Source, (www.nietzschesource.org), hosted by the Institut des Textes et Manuscrits Modernes, CNRS-ENS, Paris and at the Maison Française d'Oxford (CNRS-MAEE), Oxford, working in collaboration with specialists at the Ludwig-Maximilians-University, Munich, publishes essential and reliable scholarly editions of Nietzsche's works and Nachlass:

1. The *Digitale Faksimile Gesamtausgabe* (Digital Facsimile Edition) provides high quality reproductions of Nietzsche's published works, proofs, drafts, and preparatory manuscripts based on the original manuscripts and prints held at the Foundation of Weimar Classics. This is the largest facsimile edition ever realized within the context of Nietzsche studies. The majority of this manuscript material has never been published in facsimile and none of it has ever been published online.

2. The *Digitale Kritische Gesamtausgabe Werke und Briefe* (Digital critical edition of the complete works and letters) publishes the electronic version, encoded in TEI-P5, of the reference critical edition of Nietzsche's works, posthumous fragments and letters previously published in print by Giorgio Colli and Mazzino Montinari.

3. The *genetic edition* will give direct access to two of Nietzsche's works, *The Wanderer and His Shadow* and *Daybreak*, in a groundbreaking, experimental edition that allows to follow each step of the genesis of a text, from first sketches to its publication.

5.4 Ludwig Wittgenstein

Through *Wittgenstein Source* (www.wittgensteinsource.org), the Wittgenstein Archives at the University of Bergen at AKSIS/Unifob, Bergen, provides access to 5,000 pages of the Wittgenstein *Nachlass* in both facsimile and transcribed editions. These include typescripts and manuscripts, notebooks and fair copies, single cuttings and coherent works, including material from the Big Typescript complex (1929-1934), the Brown Book complex (1934-1936), the "Lecture on Ethics," and "Notes on Logic," in both facsimile and critical transcription, with English translation and including Wittgenstein's own translations of English texts into German and vice versa. Furthermore, *Wittgenstein Source* provides peer-reviewed scholarly articles, lectures, bibliographies.

Wittgenstein Source organizes the content into the following modules:

1) *Bergen Facsimile Edition* provides direct access to approximately 5,000 facsimiles of Wittgenstein's *Nachlass*. While these images have been published before in the *Bergen Electronic Edition* (Oxford University Press) on CD, none has been published before online, and none has been published under an Open Access licence.

2) The *Bergen Text Edition* provides, in HTML, direct access to the Diplomatic and Normalized editions of each single remark contained in the 5,000 pages of Wittgenstein's *Nachlass*. These editions are prepared on the fly by applying a stylesheet (provided by WAB), on top of the XML transcription itself.

5.5 Contemporary Philosophers - Audiovisual content

Within the Sophiavision collection, RaiNet has published a total of 300 videos complemented by abstracts and transcriptions. The main source from which contributions were selected was the Multimedia Encyclopaedia of Philosophical Science (EMSF), a collection of 1500 interview-lessons to scholars from 34 countries. In addition to a large proportion of contributions from EMSF and from some TV programs related to the Encyclopaedia (*Aforismi, Knowledge Universe*, etc.), many valuable contributions never broadcast before are included in Sophiavision, as well as completely new interviews with reference to specific texts, authors and topics provided by other partners of the Discovery project. All videos are available both in Windows Media Video and mp4.

5.6 Bibliographic description and metadata

In order to ensure the highest quality bibliographic description for all of the materials in the Philosource nodes and to maximize interoperability with aggregators and collections such as Europeana, Discovery has created its own metadata standard. Discovery's metadata schema consists of a series of Dublin Core elements (http://dublincore.org/documents/dcmi-terms/) combined with the Europeana Semantic Elements (the ESE) and is designed to exceed both standards.

The schema adheres closely to the Europeana standard with one key difference, allowing for the cataloguing of Discovery's digital critical editions. The Europeana metadata specification presupposes two scenarios: 1) the digitisation of historical materials or 2) the creation of original, born-digital works. Accordingly, contributing collections are asked to submit bibliographic metadata that describes either the historical materials themselves or the born-

digital work. Europeana's metadata specification does not allow for the creation of original works based on historical collections. The collections in Discovery contain digital critical editions based on previous works, but with a significant enough difference to be considered by scholars to be a new edition. By creating an additional date field as well as a field for the "previous publisher," works in the Philosource collections can be adequately described as original works while still providing ample credit to previous publishers, editors, and creators.

The Discovery metadata standard is intended to allow project partners to catalogue works and the smallest logical unit—a notebook, manuscript, book, etc.—in order to maximize the precision of searching. The schema, though, is also designed to allow for the recognition of those units within a larger collection. For example, Nietzsche's notepad 'N-I-3' is described as an octavo-sized notebook containing 'casual jottings' written between 1872 and 1879. The bibliographic record primarily describes the original (analog) object but contains the technical metadata required to also recognize it as also a description of the digital copy. At the same time, the record contains clear references to the facsimile of this notebook forming a part of the 'Digital Facsimile Edition.' Conversely, the record for the 'Digital Facsimile Edition' contains references to 'N-I-3' and indeed all of the parts of the Edition.
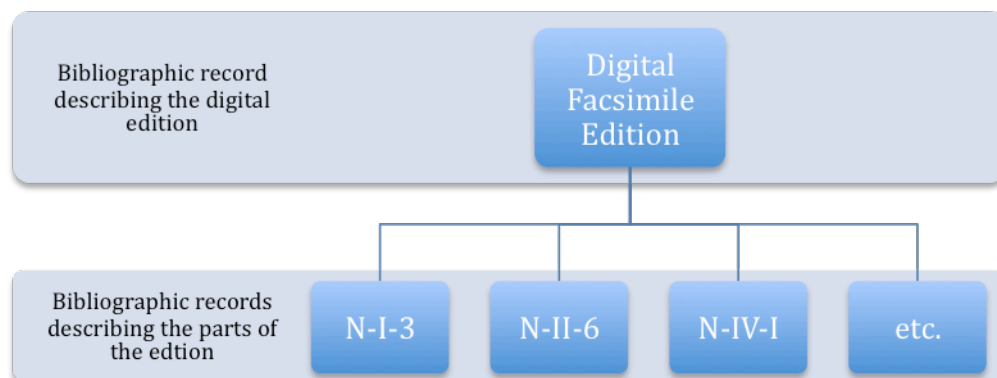
6 Impact & Sustainability

Impact

Discovery's target users are philosophers and scholars in the broader humanities, i.e. researchers, university lecturers, students in philosophy and related humanities disciplines with a special interest in the writings of Nietzsche, Wittgenstein, the Pre-Socratic's and modern philosophy in general. The goal of Discovery is first to provide *content*—that is sources for study that are extremely rare and therefore valuable—and second, to provide an easy-to-use yet technically-sophisticated infrastructure that maximizes the full potential of the internet in working with this content. Discovery's potential stakeholders are therefore, the worldwide community of philosophy scholars. If Discovery is successful it has the potential to change the way humanists go about their work as researchers, scholars, and teachers. The surveys carried out by the Oxford Internet Institute for Discovery indicate that this openness to new technologies exists to an astonishing degree among our target users.

Talia, the software developed by Discovery which powers the Philosource federation, will be made freely available to other content providers such as research groups, foundations,

libraries, and archives. The goal is a large federation of content providers that will link humanities scholars and their sources in an unprecedented way, thereby impacting the way research is carried out within a globally networked discipline. The *Philospace* desktop application will support this network by providing a state-of-the-art, socio-scientific tool that enables all of the traditional activities of a humanities scholar, i.e. collecting and adding content, annotating sources, submitting articles and commentaries for peer-review, and making work and sources available to colleagues, related research groups, and students.

Sustainability

After the end of the European funding period, the long-term sustainability will be achieved through three different strategies:

1. subsequent funding by universities and national and international funding agencies;
2. collaboration with major national libraries;
3. commercial exploitation of the technological infrastructure.

In the non-digital world, research infrastructures for the humanities such as universities and libraries, as well as scholarship, teaching, and research are conducted as a public good and therefore subsidized publicly. Digital research infrastructure ought to rely on the same type of sustainability to ensure its long-term preservation and extension.

Even if funding in the humanities and related social sciences is not of the same order of magnitude as is the support for the natural sciences and engineering, the universities and research institutions involved in the Discovery project will ensure the sustainability of the Discovery project along with their libraries. The relevance of the philosophical content produced by Discovery guarantees that the concerned institutions will at least ensure the continued existence of the Web sites and also provide human and financial resources to further enrich them. Major national libraries could also be interested in the preservation of the content.

Concerning the software part of the project, the company involved in the its development, appears to have very concrete plans for the exploitation of the developed technologies and are already well engaged in these activities, as outlined in the following section.

Exploitation of technology

The exploitation potential of the technologies developed within the context of the Discovery project—Talia and Philospace—has been investigated through participation in several conferences and trade fairs as well as through a special meeting organized by Net7 with its business partners. In particular, the technology providers of the consortium established collaboration with two Italian SMEs (SensibleLogic srl http://www.sensiblelogic.it and Hyperborea scrl http://www.hyperborea.it) who show great interest in acquiring and using the Discovery technologies Talia and Philospace. Additionally, an important collaboration has been started by Net7 with the *Centre de recherche et de restauration des musées de France* regarding the integration of the IIP technology (http://iipimage.sourceforge.net/) into Talia. These partnership-building activities have led to the joint organization of two business presentations where Net7, in partnership with other SMEs, presented and integrated technologies derived from Discovery into a business offer. In particular, 4 Italian SME's joined their products and service offerings to form a vertical consortium that offers services in the sector of technologies for cultural heritage, ranging from digital acquisition of images to cataloguing, publication on the Web, and semantic annotation. The services and products offered by this consortium have been presented during two major cultural events in Italy: *Lu.Be.C. Digital Technology 2007* and *2008, Bergamoscienza 2008* as well as at a workshop

entitled *Digitalizzare e valorizzare patrimoni culturali* organised in collaboration with important Italian cultural institutions.[1]

In September 2008, "...in the context of the French Presidency of the Council of the European Union, the European Commission, OSEO and the French Ministry for the Economy, Industry and Employment highlighted a selection of examples of successful European SMEs whose innovation projects have had a considerable economic impact due to Europe-oriented corporate strategies and to the various European support mechanisms these companies have benefited from." Net7 was selected to participate in the conference "EU Research for SMEs, Innovation in Motion" where it presented the Discovery Project and related technologies in a roundtable (a video of part of the presentation is available on the conference Website) and in a press conference[2].

During 2009, parts of the technological infrastructure developed in Discovery, have been re-used by Net7 and Hyperborea SRL to lay the foundations of BAMI (Biblioteca Aperta di Milano – Milan's Open Library), which is a semantic digital library dedicated to Italian musical heritage of the 19th century. BAMI was promoted by the Italian Ministry of Cultural Heritage, the Municipality of Milan and the Lombardia Region.

In January 2010, a large research project on cultural heritage called TemArt will be launched by the Tuscany Region. The project, whose budget is around 3M Euros, will be dedicated to bringing to the market a range of new advanced technologies for the analysis and restoration of cultural artifacts. These technologies include portable 3D microscopes, laser scanning and other advanced hardware innovations. All the data produced by the hardware tools will then be stored, semantically enriched and analyzed in a semantic digital library based on Talia. Net7, who is a key partner of the project, will reuse and improve the technology developed in Discovery to enter into and open up a new market sector.

At the time of writing this report, Net7 and UNIVPM, together with another research institution (DERI Ireland) and three other SMEs (in Poland, Malta and Italy) are in the process of submitting a proposal in the forthcoming *Research for SMEs* FP7 Capacities call. The proposal will be dedicated to the development of advanced tools for semantic digital libraries, which will complement and enrich the product range of the SMEs. If this proposal will be successful, Talia will be integrated with new Semantic Web components that will make it even more appealing for customers worldwide.

Finally, it is worth noting that thanks to their successful collaboration in Discovery, Net7 and UNIVPM have signed an agreement which, starting from January 2010, will create a new common brand for their software (Talia and Philospace). The new brand, called Muruca, will create a research, development and commercial coalition that will join together various Open Source applications that can be used as base components for Semantic Digital Libraries in the sectors of the Humanities and Cultural Heritage. In addition to Talia and Philospace, Muruca will also include Commodoro, a collaborative system for publishing XML based text editions, developed by the Scuola Normale Superiore in Pisa. The three applications will not only share a common brand, but they will also be further developed, disseminated and distributed by following a shared strategic plan. All Discovery partners will benefit from the further development of its base software infrastructure.

---

[1] See www.lubec.it, bergamoscienza.it, blog.barbz.org/wp-content/uploads/2008/07/programma_workshop.pdf.

[2] See www.eu-research-for-smes.oseo.fr.