



Knowledge Base for RTD Competencies in IST



Deliverable D5.1

First Version of the Portal with Basic Functionality

Author(s):	Jure Ferlež, Brigitte Jörg, Mitja Jermol
Identifier:	D5.1
Work package:	WP5 Portal Setup
Lead Partner:	Deutsch. Forsch. für Künstliche Intelligenz (DFKI)
Partner(s):	Institut Jozef Stefan (JSI) Ontotext Lab, Sirma Group (ONT)
State of document:	final
Version:	1.0
Dissemination Level:	Public
Date:	2005-10-23

This document is part of a SSA project funded within the IST Programme of the Commission of the European Communities – Project No: **FP6-2004-IST-3 – 015823**.

IST World Consortium

Participant Name	Participant Short Name	Country
Deutsches Forschungszentrum für Künstliche Intelligenz (Co-ordinator)	DFKI	Germany
Institute Jozef Stefan	JSI	Slovenia
Ontotext Lab, Sirma AI EAD	ONT	Bulgaria
RTD Talos	Talos	Cyprus
Institute of Information Theory and Automation	UTIA	Czech Republic
Archimedes Foundation	AF	Estonia
Computer and Automation Research Institute, Hungarian Academy of Sciences	MTA SZTAKI	Hungary
Institute of Mathematics and Computer Science, University of Latvia	IMCS	Latvia
Lithuanian Innovation Centre	LIC	Lithuania
Projects in Motion	PiM	Malta
Technical University of Silesia	SUT	Poland
National Institute for Research and Development in Informatics	ICI	Romania
Silesian University of Technology	STUBA	Slovakia
TUBITAK	TUB	Turkey
CCLRC	CCLRC	United Kingdom

Abstract

The IST World project aims at setting up and populating an information portal with innovative functionalities that helps to promote RTD competencies in IST on a local, national and European level. In order to acquire as early end user feedback of the functionality and usability, and to support the dissemination process, the IST World project has adopted a cyclic development strategy. We have setup the first version of the IST World portal which contains a critical mass of data and which is primarily intended for testing by project partners and subsequently by the User Steering Board. The first version of the portal available at <http://www.ist-world.org/> provides Web access to the new integrated IST World data repository. The data of three distinct data sources SICRIS, CORDIS – FP6 and a subset of GOOGLE SCHOLAR was acquired, cleaned and integrated into the IST World data store.

Supported functionalities of the current portal are:

- Full text constraint searching and browsing of the data held in the integrated database.
- Advanced visual analytic technique of collaboration diagram on collaboration information between the different entities represented in the information store.
- Import / Export interface for CERIF-compliant XML datasets

Content Table

Knowledge Base for RTD Competencies in IST	1
IST World Consortium	2
Abstract	3
Content Table	4
1. Introduction	5
2. Information Repository	5
3. Portal Functionality	6
3.1. Browse	7
3.2. Analyze.....	9
3.3. Forecasting and Prediction	14
3.4. For Members	14
3.5. Import / Export Interface for CERIF-compliant XML datasets.....	15
4. Additional Work	15
4.1. Content / Data	15
4.2. Portal Architecture.....	15
4.3. Portal Functionalities.....	15
5. Conclusion	16
6. Bibliography.....	16

1. Introduction

The IST World project aims at setting up and populating an information portal with innovative functionalities that helps to promote RTD competencies in IST on a local, national and European level. In order to acquire as early end user feedback of the functionality and usability, and to support the dissemination process the IST World project has adopted a cyclic development strategy. We have setup the first version of the IST World portal which contains a critical mass of data and which is primarily intended for testing by project partners and subsequently by the User Steering Board. The First version of the portal available at <http://www.ist-world.org/> provides Web access to the new integrated IST World data repository. The data of three distinct data sources SICRIS, CORDIS – FP6 and a subset of GOOGLE SCHOLAR was acquired cleaned and integrated into the IST World data store.

Supported functionalities of the current portal are:

- Full text constraint searching and browsing of the data held in the integrated database
- Advanced visual analytic technique of collaboration diagram based on collaboration information between the different entities represented in the information store.
- Import / Export interface for CERIF-compliant XML datasets

This document is organized as follows: Section two describes the realized IST World data repository. Section three explains the functionalities of the first version of the IST World Web portal. Section four is about additional work.

2. Information Repository

The IST World information repository is build with the MS SQL Server RDMS system provided by Microsoft. The MS SQL Server 2005 is installed on a Windows 2003 Server edition operation system that runs on a fast 64 bit computer designated solely for the task of serving IST World community requests.

The database schema employed follows the CERIF methodology by extending the base CERIF relational model as described in deliverables *Definition of Central Data Structure* (D1.1) and *Data Model for Knowledge Organization* (D1.2). The database creation and the application of the schema were realized by implementing SQL scripts provided by the CERIF task group and extended with additional schema entities defined in D1.1. The scripts needed to be adjusted to the specifics of the SQL interpreter provided by the MS SQL Server system. The installed database schema for the IST World repository is quite complex and consists of 361 tables.

The data contained in the first version of the information repository is built on three different data sets:

D5.1: First Version of the Portal with Basic Functionality

- **The SICRIS data:** This CERIF-reformulated CRIS database encloses de-facto information about all Slovenian research activities. There are 579 research organizations, 11147 researchers, 2982 research projects, 597 research programs instances in the database describing researchers, research organizations and the research projects in Slovenia that are supported by the Ministry of Education and Sport and the Ministry of Science and Higher Education
- **GOOGLE SCHOLAR data:** Google Scholar data contains information about scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from most of the areas of research. It uses Web crawling technique to find out publications from a wide variety of academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles that are accessible on the Web. We have used Web crawling mechanisms to automatically acquire a subset of items published by the Google Scholar Web site. The subset contains 63474 persons and 58300 publications.
- **Project Intelligence (PI) – CORDIS data:** The 6th Framework Program project database was automatically built from the data that is publicly available on the CORDIS information system and from internal data stores of the European Commission. This data was effectively used by the Project Intelligence portal [1]. The subset contains 3181 organizations and 663 projects.

All together there are approximately 141.000 entities available in the database. The LT World data are available in CERIF-compliant XML format according to the schema specification of deliverable 3.1 and will be imported into the IST World repository in the next step.

3. Portal Functionality

The main goal of the IST World portal is to integrate data about European research activities, improve the process of getting insight into activities and provide means for computer aided social networking. The first version of the portal provides search over the integrated database and subsequently an analysis of the search results by means of a collaboration diagram. These functionalities are adopted from the already existing functionalities of the Project Intelligence and the LT World portal.

The current technical interface (back end) of the IST World repository functionally allows for a simple and efficient integration of CERIF-compliant XML datasets from different data sources as will be described in section 3.5. The graphical user interface (front end) of the IST World portal represents IST World functionalities as can be seen in the top menu in figure 1.




Figure 1: IST World top-level menu.

The first version of the Web portal is realized using Microsoft’s Internet Information Server based on the latest Microsoft .NET v2.0 implementation framework. The Web page design follows the ASP.NET programming framework recommendations.

3.1. Browse

A complex and constraint enabled full text search engine allows for searching the IST World information database by keywords. The result page based on a simple query for “germany” within Organisations is displayed in Figure 2.



The screenshot shows the IST World portal interface. At the top left is the 'istworld' logo. To the right, there is a link: 'Join the IST World community >>>> Learn how to use the portal >>>'. Below this is a navigation bar with tabs: 'BROWSE', 'ANALYZE', 'FORECASTING & PREDICTION', 'FOR MEMBERS', and 'ABOUT PROJECT'. The 'BROWSE' tab is active.

On the left side, there is a 'BROWSE:' section with a list of categories: 'Organizations', 'Projects', 'Experts', and 'Publications'. Below this is an 'Organizations:' section with a search input field containing 'germany', a 'Simple search' button, and a link to 'Advanced search'.

The main content area displays the search results. It starts with the text: 'There are 422 organizations matching your search.' Below this is a list of organization names, each separated by a dashed line. The list includes: 4M2C PATRIC SALOMON GMBH, ABSINT ANGEWANDTE INFORMATIK GMBH, ACIT ADVANCE CONCEPTS FOR INTERACTIVE TECHNOLOGY GMBH, ADAM OPEL AKTIENGESELLSCHAFT, ADVANCED MASK TECHNOLOGY CENTER GMBH & CO KG, ADVANCED REALTIME TRACKING GMBH, AIRBUS DEUTSCHLAND GMBH, AIXTRON AG, ALBERT LUDWIGS UNIVERSITAET FREIBURG, ALCATEL SEL AG, ALLERTON INTERWORKS COMPUTER AUTOMATION SYSTEMS GMBH, ALLIANZ ZENTRUM FUER TECHNIK GMBH, ALTICAST GMBH, ARNOLD & RICHTER CINE TECHNIK GMBH & CO BETRIEBS KG, ART + COM MEDIEN TECHNOLOGIE UND GESTALTUNG AKTIENGESELLSCHAFT, ATMEL GERMANY GMBH, AUDI AG, AUDI ELECTRONICS VENTURE GMBH, AUTOBAHNDIREKTION SUEDBAYERN, and AVE VERKEHRS UND INFORMATIONSTECHNIK GMBH. At the bottom of the list are the numbers '1 2 3 4 5 6 7 8 9 10 ...'.

On the right side, there is an 'ANALYZE:' section with a list of options: 'Collaboration Diagram', 'Competence Diagram (Second version available in February 2006)', and 'Group Discovery (Second version available in February 2006)'.

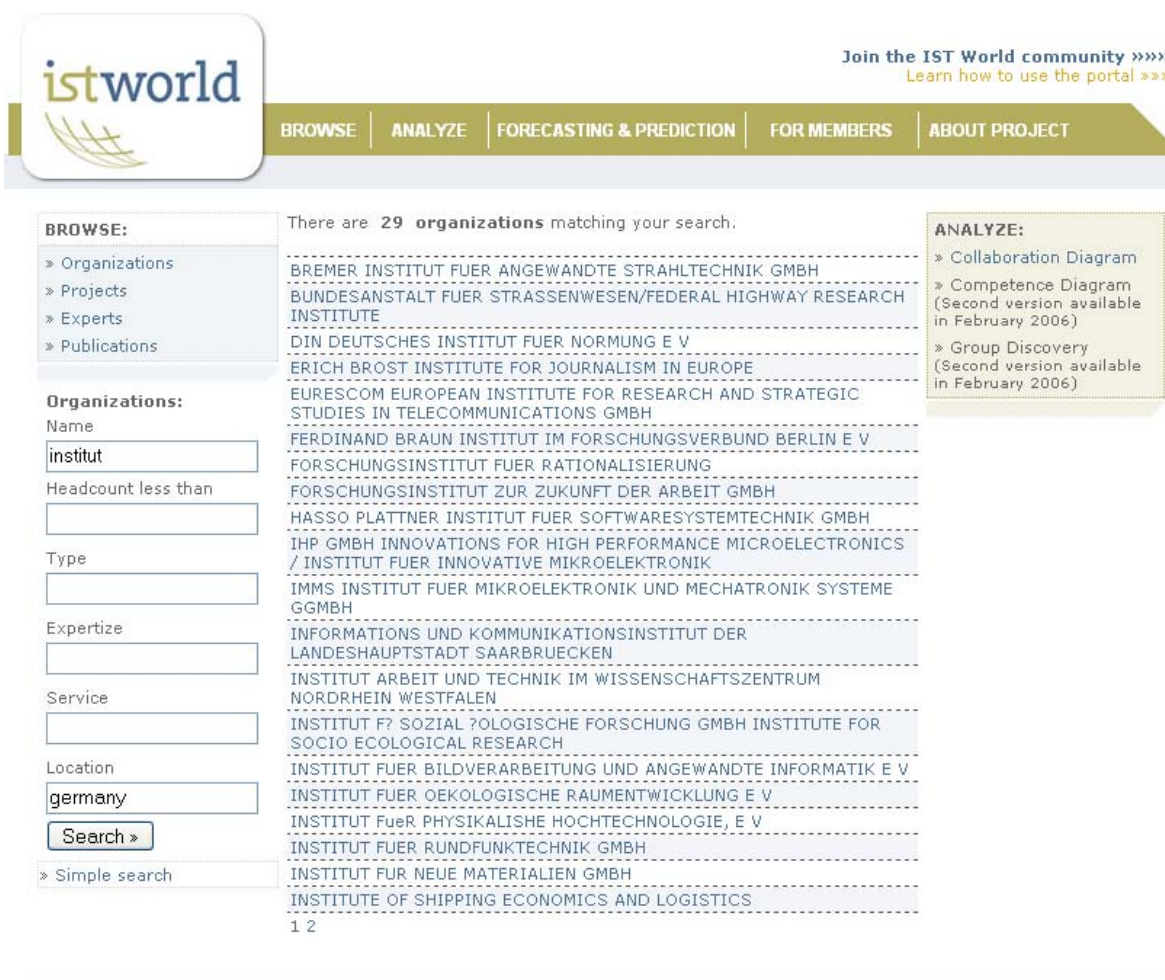
At the bottom of the page, there is a footer: 'Please send corrections and pointers to missing information to feedback@ist-world.org.' Below this is the logo for 'Information Society Technologies'.

Figure 2: Simple search for all organizations related to Germany.

The BROWSE functionality by default enables a simple full text search of organizations with respect to their expertise and location as shown in figure 2.

A more advanced search template is displayed in figure 3. Within the advanced view users can query for organizations according to specified conjunctive constraints. Additional to results for organizations related to Germany in figure 2, the advanced search allows the refinement of the previous query with constraints on specified fields like Name and Location. The results displayed in figure 3 show organizations that have institut in their name **and** are located in Germany

D5.1: First Version of the Portal with Basic Functionality



The screenshot shows the IST World portal interface. At the top left is the 'istworld' logo. To the right, there is a link: 'Join the IST World community >>>> Learn how to use the portal >>>'. Below this is a navigation bar with buttons for 'BROWSE', 'ANALYZE', 'FORECASTING & PREDICTION', 'FOR MEMBERS', and 'ABOUT PROJECT'. The main content area is divided into three sections: 'BROWSE:', 'ANALYZE:', and a central list of search results. The 'BROWSE:' section has a dropdown menu with 'Organizations' selected. The 'ANALYZE:' section has a dropdown menu with 'Collaboration Diagram' selected. The central list shows 29 organizations matching the search criteria, with the first few being: BREMER INSTITUT FUER ANGEWANDTE STRAHLTECHNIK GMBH, BUNDESANSTALT FUER STRASSENWESEN/FEDERAL HIGHWAY RESEARCH INSTITUTE, DIN DEUTSCHES INSTITUT FUER NORMUNG E V, ERICH BROST INSTITUTE FOR JOURNALISM IN EUROPE, EURESCOM EUROPEAN INSTITUTE FOR RESEARCH AND STRATEGIC STUDIES IN TELECOMMUNICATIONS GMBH, FERDINAND BRAUN INSTITUT IM FORSCHUNGSVERBUND BERLIN E V, FORSCHUNGSINSTITUT FUER RATIONALISIERUNG, FORSCHUNGSINSTITUT ZUR ZUKUNFT DER ARBEIT GMBH, HASSO PLATTNER INSTITUT FUER SOFTWARESYSTEMTECHNIK GMBH, IHP GMBH INNOVATIONS FOR HIGH PERFORMANCE MICROELECTRONICS / INSTITUT FUER INNOVATIVE MIKROELEKTRONIK, IMMS INSTITUT FUER MIKROELEKTRONIK UND MECHATRONIK SYSTEME GGMBH, INFORMATIONEN UND KOMMUNIKATIONSINSTITUT DER LANDESHAUPTSTADT SAARBRUECKEN, INSTITUT ARBEIT UND TECHNIK IM WISSENSCHAFTSZENTRUM NORDRHEIN WESTFALEN, INSTITUT F?R SOZIAL ?OLOGISCHE FORSCHUNG GMBH INSTITUTE FOR SOCIO ECOLOGICAL RESEARCH, INSTITUT FUER BILDVERARBEITUNG UND ANGEWANDTE INFORMATIK E V, INSTITUT FUER OEKOLOGISCHE RAUMENTWICKLUNG E V, INSTITUT FUER PHYSIKALISCHE HOCHTECHNOLOGIE, E V, INSTITUT FUER RUNDFUNKTECHNIK GMBH, INSTITUT FUR NEUE MATERIALIEN GMBH, and INSTITUTE OF SHIPPING ECONOMICS AND LOGISTICS. There are pagination links '1 2' at the bottom of the list.

Please send corrections and pointers to missing information to feedback@ist-world.org.



Figure 3: Advanced search for institutes in Germany

The results of the queries are presented in alphabetical order and enable navigation to the details of specific search results via hyperlink. By clicking on one of the results from that list the detailed view of the entity can be explored. The IST World repository holds the entities Project, Person, Organization and Publication, which are to be queried in a similar way.

Figure 4 presents a detailed view of one individual project entity, contained in the IST World information store indicating relations to organizations by hyperlinks at the bottom of the projects' view. One can browse from within the project instance to partner organization instances involved in the project and so on. Browsing from entity to entity is enabled on the basis of collaboration data. This functionality adapts the LT World and Project Intelligence navigation functionality between entities based on entity references.



BROWSE | ANALYZE | FORECASTING & PREDICTION | FOR MEMBERS | ABOUT PROJECT

Title
A Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information systems

Keywords
 Information Processing, Information Systems; Life Sciences; Medicine, Health; Scientific Research

Abstract
 The objective of the ARTEMIS project is to develop a semantic Web services based interoperability framework for the health care domain. We take a highly innovative approach regarding the interoperability of medical information systems with respect to the current approaches. We focus on processes in terms of Web services rather than recording and documentation of electronic health records. In other words, our approach allows a standard way of accessing the data since there are very many standards that need to work together. The interoperability problems of medical information systems are two fold: First there are multiple, incompatible, proprietary approaches to connecting disparate applications. Secondly, there are more than one standard to represent the same information, which in turn creates an interoperability problem. ARTEMIS will enable medical practitioners to access patient records securely, seamlessly through a low-cost peer-to-peer infrastructure, regardless of where their patients or their records might be. ARTEMIS project provides the healthcare industry with an ideal platform to achieve difficult integration problems. Our Web service model encapsulates already existing applications and access to documents in a standard way and incorporates service providers, service consumers and service registries. Currently most prominent Web service registries are Universal Description, Discovery, Integration (UDDI) and electronic business XML (ebXML). There are also very recent efforts to use Peer-to-peer networks based on Web services. However both service registries and P2P architectures available do not provide semantically enriched search capabilities. In the ARTEMIS project we will provide extensions to these architectures to enable discovery of the Web services based on their semantic descriptions. Medicine is one of the few domains to have some domain knowledge in a computable form, which we will exploit in defining the semantics of medical Web services.

Collaborative Projects
 |

Resulting Publication
 |

People

Last Name	First Name	Role

Organizations

ALTEC INFORMATION AND COMMUNICATION SYSTEMS S A	Partner
KURATORIUM OFFIS E V	Partner
MIDDLE EAST TECHNICAL UNIVERSITY	PrimePartner
SOUTH AND EAST BELFAST HEALTH AND SOCIAL SERVICES TRUST	Partner
TEPE TEKNOLOJIK SERVISLER ANONIM SIRKETI	Partner
THE UNIVERSITY OF SOUTHAMPTON	Partner

Funding Programmes

Name	Budget	Currency	Role
eHealth	1990000	EURO	unknown

Figure 4: Project details information view.

3.2. Analyze

The automated analysis process consists of two steps. First, in the *Data Selection* step, the user selects the entities or a subset of entities that represent the interest or target of the user analysis. In the second *Analysis Specification* step the user selects one of the provided analytic methods, together with specifying the necessary parameters to define the direction of the analysis process. The *Data Selection* step is realized by re-using the BROWSE functionality what enables the user to re-use any subset of data for analytical processing. This setting also supports a quick and effective implementation of additional visualization techniques on top of the *Data Selection* step.

D5.1: First Version of the Portal with Basic Functionality

In the first version of the portal the Collaboration Diagram visualization analysis technique is implemented in order to enable an analysis of search results. It supports a computation and the visualization of different types of collaboration diagrams dependant on display and query definitions. Entities are presented as vertexes and collaboration count between entities as edges. The more there are collaborations between two entities, the bolder is the displayed edge as can be seen in the following figure 5.

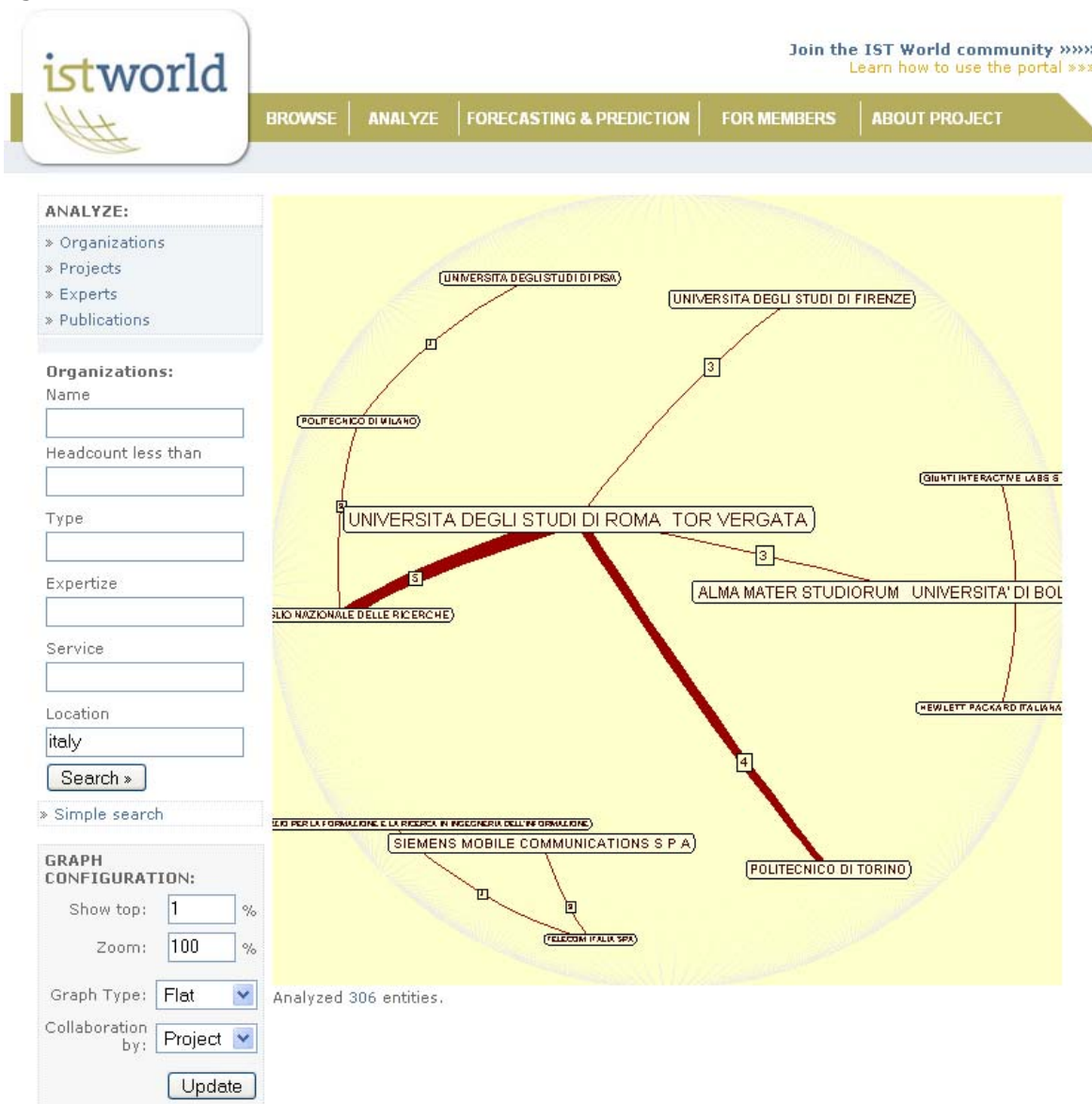


Figure 5: Visualization of collaboration results between organizations in Italy.

The current version of the collaboration diagram enables a visualization of collaboration between organizations, people, projects and publications. Two organizations are considered to be collaborating if they take part in a common project (collaboration by Project) or if there are people affiliated with both organizations (collaboration by Person) or they published any joint paper (collaboration by Publication).

D5.1: First Version of the Portal with Basic Functionality

An example of a collaboration diagram has been introduced in Figure 5. The ANALYZE functionality enables a simple selection of data for collaboration analysis (*Data Selection*) by re-using the BROWSE functionality and moreover allows graph configuration by defining visualization parameters (*Analysis Specification*) as displayed in Figure 5. The GUI of the portal supports the usage of the analysis tools.

A possible scenario of a *Data Selection* and *Analysis Specification*: A user wants to find the most collaborative institutions by project in Italy.

1. User chooses the *BROWSE* functionality to obtain a list of all organizations located in Italy as displayed in Figure 6.
2. User clicks the right hand side hyperlink *Collaboration Diagram* to analyze and visualize the collaboration between institutions in Italy as displayed in Figure 7.
3. User configures graph visualization by setting the parameter *Show top 1%* to select only the top 1 percent of all collaborations as displayed in Figure 8.
4. User re-uses the *Data Selection* step by applying top left hand side search GUI to narrow the object of analysis only to organizations with word "university" in their name. A new collaboration diagram is then generated as shown in Figure 9.



The screenshot shows the istworld portal interface. At the top left is the istworld logo. To the right, there is a link: "Join the IST World community >>>> Learn how to use the portal >>>". Below this is a navigation bar with tabs: BROWSE, ANALYZE, FORECASTING & PREDICTION, FOR MEMBERS, and ABOUT PROJECT. The main content area is divided into three sections:

- BROWSE:** A sidebar menu with options: Organizations, Projects, Experts, and Publications.
- Organizations:** A search form with fields for Name, Headcount less than, Type, Expertize, Service, and Location (set to 'italy'). A "Search >" button is at the bottom.
- Search Results:** A list of 306 organizations. The first few visible are: 3VSOFT S R L, 4SITE SRL, 5T S C R L, A C NIELSEN ITALIA SPA, ABDUS SALAM INTERNATIONAL CENTER FOR THEORETICAL PHYSICS, ADVANCED COMPUTER SYSTEMS A C S S P A, AEROPORTI DI ROMA SPA, AGENZIA SANITARIA REGIONALE EMILIA ROMAGNA, ALCATEL ITALIA S P A, ALENIA AERONAUTICA SPA, ALENIA MARCONI SYSTEMS S P A, ALENIA SPAZIO S P A, ALESSI S P A, ALMA MATER STUDIORUM UNIVERSITA' DI BOLOGNA, AMITIE SRL, ARAKNE S R L, ARCADIA DESIGN S R L, ASSOCIAZIONE DEI FONOGRAFICI ITALIANI, ASSOCIAZIONE IMPRESA POLITECNICO, and ASSOCIAZIONE ITALIANA PER LA RICERCA INDUSTRIALE. A pagination bar shows "1 2 3 4 5 6 7 8 9 10 ...".
- ANALYZE:** A sidebar menu with options: Collaboration Diagram, Competence Diagram (Second version available in February 2006), and Group Discovery (Second version available in February 2006).

Please send corrections and pointers to missing information to feedback@ist-world.org.



Figure 6: Query definitions for all organizations located in Italy.



Figure 7: User selected to visualize the collaboration by projects of all organizations in Italy.

D5.1: First Version of the Portal with Basic Functionality

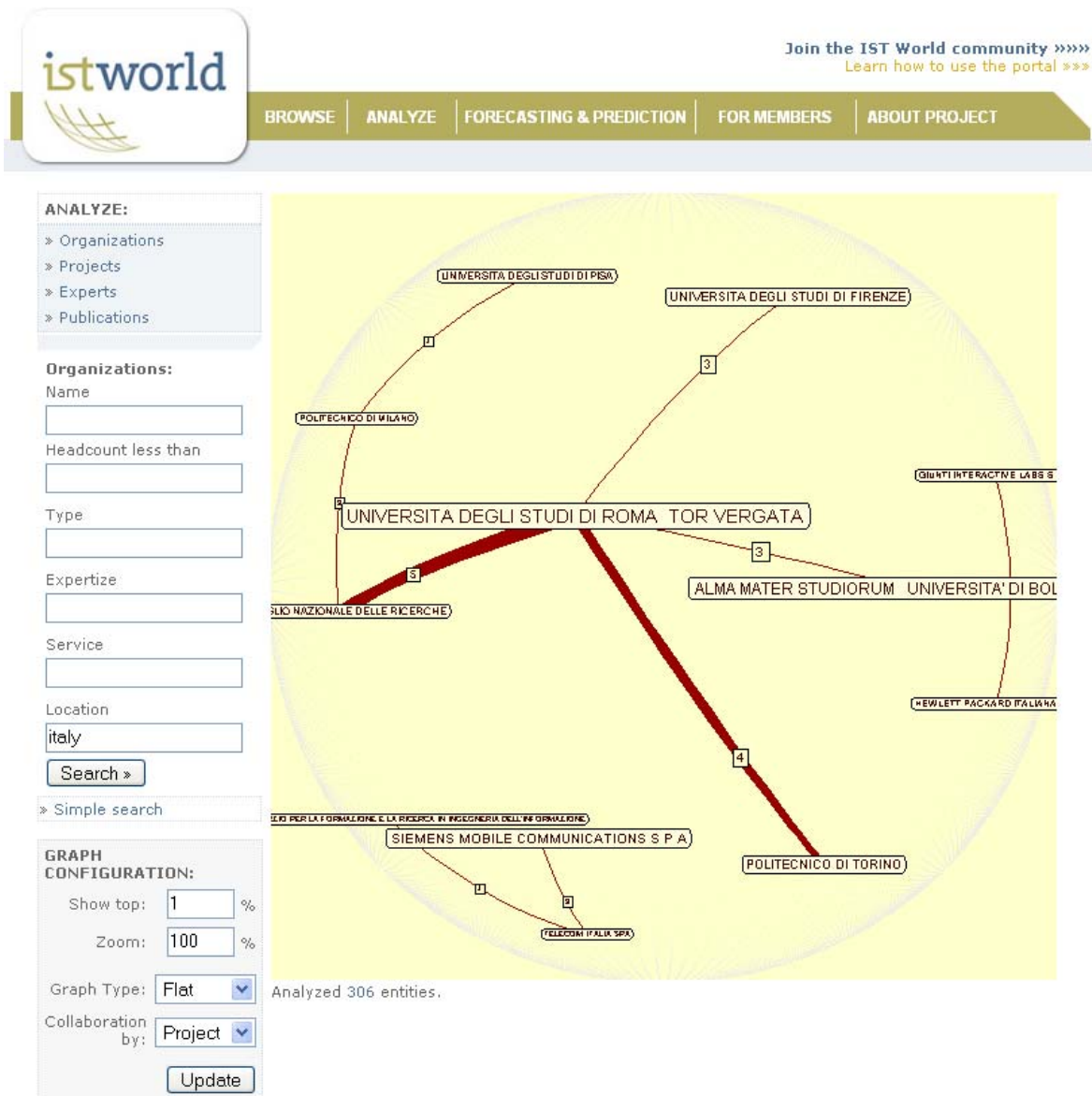


Figure 8: View only top 1% of collaboration by projects of all the organizations in Italy.

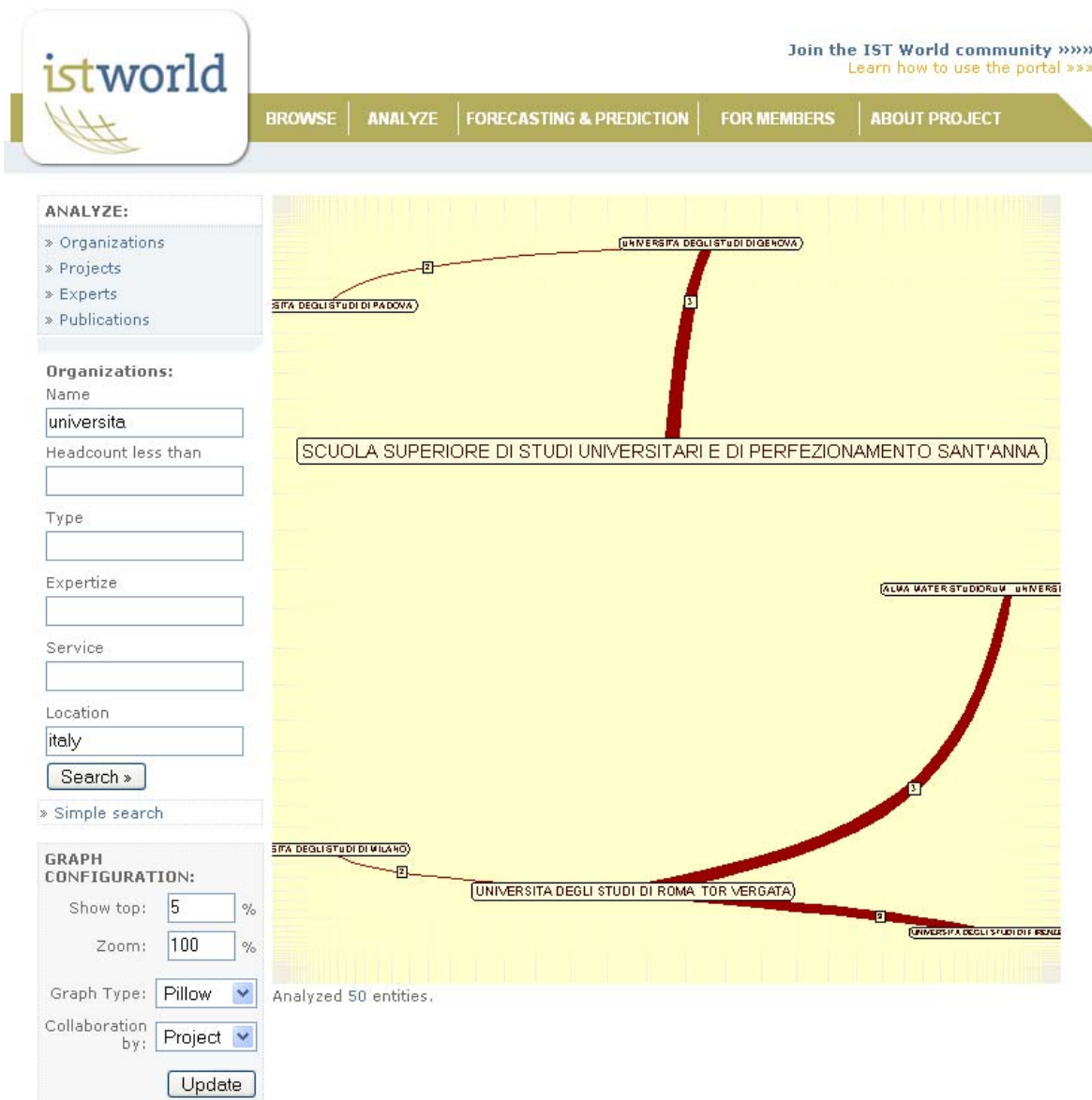


Figure 9: Narrow the analysis object from general organizations to only universities in Italy and only show top 5% of collaborations using the hyperbolic distortion (pillow) graph type.

3.3. Forecasting and Prediction

According to the Description of Work the Forecasting and Prediction functionality is subject of deliverables 5.2 Portal with Enhanced Functionality and 5.3 Portal with Full Functionality.

3.4. For Members

According to the Description of Work For Members functionality is subject of deliverables 5.2 Portal with Enhanced Functionality and 5.3 Portal with Full Functionality.

3.5. Import / Export Interface for CERIF-compliant XML datasets

An Import / Export interface for CERIF-compliant XML datasets was implemented in order to provide simple and efficient means for data integration. The interface is currently implemented as a standalone program, which was in part automatically generated using the Altova MapForce utility. The utility used the CERIF schema definition of the IST World repository and CERIF compliant XML schemas as defined in the deliverable 3.1 *Data Import and Export Specification as XML Schema* to in part automatically generate a mapping program, which allows quick data transformations between different storage formats.

The Import / Export interface for CERIF-compliant XML datasets thus enables simple and efficient integration of data from different data sources, given they are presented in CERIF compliant XML datasets. This will allow a simple and fast population of the IST World data repository.

4. Additional Work

4.1. Content / Data

CERIF-compliant XML data from LT World are prepared and will be imported in the next steps.

Sample data from all participating partners have been requested and were analyzed in Deliverable 3.2 *Base set of Data*.

The real national datasets have to be prepared for CERIF-compliant imports. Discussions for individual national data specifications are on the agenda of the 1st IST World consortium meeting in Vilnius, Lithuania by the end of this month. The deliverable 3.2 will serve as a means for starting discussions on the national datasets.

4.2. Portal Architecture

Content Classification. A classification scheme will be developed as part of deliverable D2.3 Portal Architecture Specification, as described in deliverable D1.2 Data Model for Knowledge Organization. This will improve navigation and visualization functionalities and moreover content classification of entities. Subsequently it will improve search results based on thematic queries.

4.3. Portal Functionalities

Web forms for manual online data imports will be implemented based on the Import/Export interface for CERIF-compliant XML datasets described in section 3.5.

To adequately implement a Web form interface usable by all participating partners first the national datasets have to be further analyzed. The analysis of the national datasets has been started with deliverable 3.2 and will be continued in discussions among all the partners at the first Consortium meeting in Vilnius.

5. Conclusion

The first version of the portal has been successfully implemented. It is set up with an acquisition of data from three different sources, which were cleaned and integrated. This was done according to the methodology defined in deliverable D1.1 Definition of Central Data Structure. The portals' web site was established (<http://www.ist-world.org/>) and a GUI was implemented. The provided functionality currently enables searching, browsing some types of visual analysis of the data that is held in the information repository. The presented first version of the portal is now ready for testing and usability analysis. First, it will be spread among the IST World project partners for immediate feedback. Next, the User Steering Board will be contacted for concrete analysis of the first portal version. The received feedback will then be used when setting up the next versions of the portal with enhanced and additional functionalities according to the cyclic application development paradigm. The portals' functionalities undergo continuous improvement processes during the whole phase of the IST World project.

IST World dissemination activity heavily relies on an early live portal presentation. The first version of the portal supports this by providing a first glance of the final state-of-the-art analysis techniques of the research related information and moreover with a critical mass of data inherent in the IST World portal.

6. Bibliography

[1] Grobelnik, M., Mladenic, D., (2002) Analysis of IT projects funded by the European Commission in 5FP. Technical Report IJS-DP 8678, Jozef Stefan Institute, Ljubljana, Slovenija, Nov. 2002.