Information and Communication Technologies

EPIWORK

Developing the Framework for an Epidemic Forecast Infrastructure

http://www.epiwork.eu

Project no. 231807

D 3.3 Public Release of the Epidemic Marketplace Platform

Period covered:
Start date of project: February 1st, 2009
Due date of deliverable: September 30th 2010
Distribution: Public

Date of preparation:
September 30th 2010
Duration:
Actual submission date:
September 30th 2010
Status:

Project Coordinator: Alessandro Vespignani
Project Coordinator Organisation Name: ISI Foundation
Lead contractor for this deliverable: FFCUL

Work package participants

The following partners have taken active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

Mário J. Silva, Francisco M. Couto, Dulce Domingos, Juliana Duque, Hugo Ferreira, Luís F. Lopes, Daniela Paolotti, Fabrício Silva, Patrícia Sousa, João Zamite.

We thank the members of the GLEaM research team who contributed with relevant datasets that we will use in demonstrations in subsequent phases of the project.

Change log

Version	Date	Amended by	Changes
0.1	2010-09-15	Mário J. Silva	First draft
1.0	2010-09-30	Mário J. Silva	Publication
1.1	2011-03-15	Mário J. SIlva	Fixed Refs.

D 3.3 Public Release of the Epidemic Marketplace Platform

Mário J. Silva[1], Francisco M. Couto[1], Dulce Domingos[1], Juliana Duque[1], Hugo Ferreira[1],
Luís F. Lopes[1], Daniela Paolotti[2], Fabrício Silva[1], Patrícia Sousa[1], João Zamite[1]

LASIGE, University of Lisbon[1], Portugal and ISI[2], Italy
epiwork@di.fc.ul.pt

Introduction

This report describes the architecture and deployment status of the Epidemic Marketplace (EM) at the time of its release to the scientific community at large, by the end of Month 20 of the Epiwork project. This constitutes release 1.0 of the EM.

In addition to new versions of its initial components and new curated contents, the EM now offers a uniform user interface and comprehensive editing capabilities for adding meta-data to the resources deposited into the Repository. Improvements to the back-office and service monitoring capabilities added increased reliability to the platform as its user base is enlarged.

The Epidemic Marketplace is available at http://www.epimarketplace.net/. The contents that have been made available to the epidemic community at the time of release of this resource have been annotated by the authors and obtained from various sources. We are thankful to the members of the GLEaM research team who contributed with relevant datasets . We will use these datasets for demonstrating the EM in subsequent phases of the project.

Epidemic Marketplace Architecture and Software Implementation status

The Epidemic Marketplace can be defined as a *distributed virtual repository*, a platform supporting transparent, seamless access to distributed, heterogeneous and redundant resources [10]. It is a virtual repository because data can be stored in systems that are external to the Epidemic Marketplace, and it provides transparent access because several heterogeneities are hidden from its users.

The Epidemic Marketplace can be composed of a set of interconnected data management nodes geographically distributed, sharing common canonical data models, authorization infrastructure and access interfaces. Data can be either stored in one or more repositories or retrieved from external data sources using authorization credentials provided by clients. Data can also be replicated among repositories to improve access time, availability and fault

tolerance. However, data replication is not mandatory; in several cases data must be stored in a single site due to, for instance, privacy constraints. It is worth noting, though, that any individual repository that composes the Marketplace will enable virtualized access to these data, once a user provides adequate security credentials.

An Epidemic Marketplace node can be composed by the following modules:

Repository: stores epidemic data sets and epidemic ontologies to characterise the semantic information of the data sets;

Mediator: a collection of web services that will provide access to internal data and external sources, based on a catalogue describing existing epidemic databases through their meta-data using state-of-the-art semantic-web/grid technologies;

Collector: retrieves information of real-time disease incidences from publicly available data sources, such as social networks; after retrieval, the collector groups the incidences by subject and creates data sets to store in the repository;

Forum: allows users to post comments on integrated data from other modules, fostering collaboration among modellers;

Several open-source tools and open standards are being used in EM 1.0, forming a software layer between the base software at each node and the epidemic applications facing the users:

- We selected **Fedora Commons** [7] and **Muradora** [14] for the implementation of the main features of the repository.

- Access control in the repository is based on the **XACML** [16], **LDAP** [24] and **Shibolleth** [19] standards (Shibolleth is still not operational at the time of release to the public).

The implementation of the mediator services started at the time of the release of the prototype to the consortium [21] (September 2009) and an initial release is now deployed [20]. The initially offered services include:

Search Objects: keyword-based client searches on the resources in the repository.

Search Collections: client searches on collections (groups of resources) in the repository.

Fetch: client retrieval of resources from the repository.

Upload: clients directly add resources and their meta-data descriptions to the repository.

The implementation of deployed mediator services follows OAI standards [15], such as ORE and PMH, and expose methods that invoke the EM deployed components. EM Clients may search and query datasets and corresponding metadata through a RESTfull interface, after an initial authentication procedure.

A new version of the Collector services, MEDCollector [27][28], has also been implemented and deployed to the EM production environment, since the release of the first prototype to the consortium in September 2009. The new version has a graphical user interface and the capability of collecting data both actively and passively from multiples sources. EM administrators can dynamically configure new data collection processes thorough the graphical interface and a number of pre-defined services.

Meta-data creation has also been significantly revised. A new meta-data editor, which is invoked when a user decides to upload a new dataset to the Repository now requires less human intervention, as many of the meta-data fields are automatically completed following the EPIWORK guidelines [9], and only leaving the task of confirming the meta-data to the user [8][10].

The Forum is implemented using phpBB [17] and is now integrated through the EM linkage structure with other modules of the Epidemic Marketplace, as when initially released in September 2009.

We have published the application-specific software developed for the EM in a Google Code [3] source repository, which is under versioning through a Subversion 1.6 server.

Deployment

Lisbon was the first site where an Epidemic Marketplace node has been deployed. A first prototype has been operational since Month 12, open for consortium use [21]. The release to the public signals the start of a much more stable and robust version, which we have been improving as we began to use it and observe how our internal users interact with it.

Hardware – We are running on the same hardware that was setup at the beginning of the project and configured for hosting the prototype released to the Consortium:

• Two DELL servers, with two quad-core processors (Athlon), 16GB of main memory
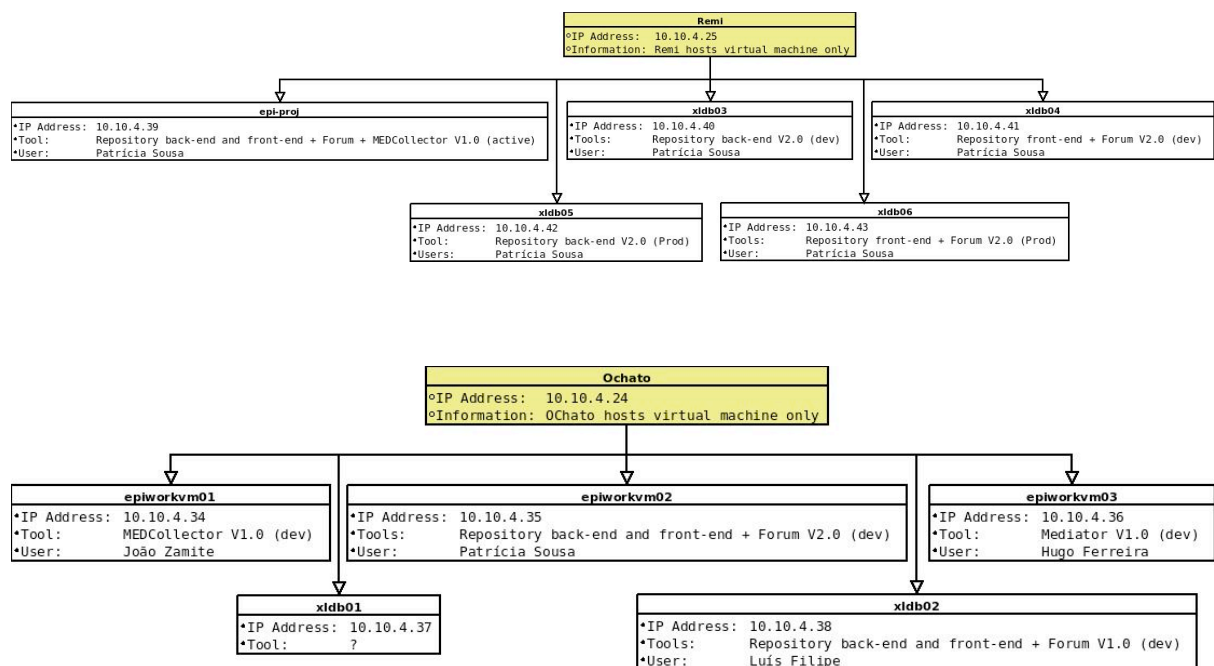
and two 1TB disks, each.

- Two Iomega network storage units, with four 1TB disks in a RAID 5 configuration.

- Computing servers and network storage units are interconnected through 1Gbit Ethernet links.

Base Software – We have completely reorganized the base configuration of the system. Both DELL servers now have an installed virtualization layer, through the Xen open source platform [23][26]. Each DELL server now hosts several virtual machines, each running a version of Linux, and dedicated to supporting primarily one of the modules described in the previous section.

In addition, have created two separate environments, one for the production version that is now accessible to the public and an internal development environment, where the next version of the EM is now being developed.

The diagrams below illustrate the base configuration of the system, showing the use/software that is assigned to each virtual machine at the two servers (named *remi* and *ochato*):

Remi	
◦IP Address: 10.10.4.25	
◦Information: Remi hosts virtual machine only	

epi-proj	
•IP Address: 10.10.4.39	
•Tool: Repository back-end and front-end + Forum + MEDCollector V1.0 (active)	
•User: Patrícia Sousa	

xldb03	
•IP Address: 10.10.4.40	
•Tools: Repository back-end V2.0 (dev)	
•User: Patrícia Sousa	

xldb04	
•IP Address: 10.10.4.41	
•Tool: Repository front-end + Forum V2.0 (dev)	
•User: Patrícia Sousa	

xldb05	
•IP Address: 10.10.4.42	
•Tool: Repository back-end V2.0 (Prod)	
•Users: Patrícia Sousa	

xldb06	
•IP Address: 10.10.4.43	
•Tools: Repository front-end + Forum V2.0 (Prod)	
•User: Patrícia Sousa	

Ochato	
◦IP Address: 10.10.4.24	
◦Information: OChato hosts virtual machine only	

epiworkvm01	
•IP Address: 10.10.4.34	
•Tool: MEDCollector V1.0 (dev)	
•User: João Zamite	

epiworkvm02	
•IP Address: 10.10.4.35	
•Tools: Repository back-end and front-end + Forum V2.0 (dev)	
•User: Patrícia Sousa	

epiworkvm03	
•IP Address: 10.10.4.36	
•Tool: Mediator V1.0 (dev)	
•User: Hugo Ferreira	

xldb01	
•IP Address: 10.10.4.37	
•Tool: ?	

xldb02	
•IP Address: 10.10.4.38	
•Tools: Repository back-end and front-end + Forum V1.0 (dev)	
•User: Luís Filipe	

Backup, Logging and Monitoring

As part of the work for preparing the platform for public release, we took extra care in setting-up and verifying backup and recovery procedures for the virtual machines on top of

the data replication architecture that we deployed earlier, where epidemic data is mirrored in the two Iomega storage units (which in turn replicate local data under a RAID configuration).

All the EM services are under monitoring from an internal Nagios platform [13], which automatically alerts the EM support in case of failure of one of the EM components.

In addition to keeping the log files of the web and middleware servers for further analysis, we also have the EM public website monitored by the Google Analytics platform [4], for access statistics data collection.

User Access

Access to many of the EM resources requires previous registration and is granted to authenticated users only. On October 1st we began accepting external registrations. Anyone who is not registered cannot insert new data, but may enter the repository and browse the collections with permissions granting access to the public

Datasets

We have continued adding resources to the EM repository since it was first released internally, not only to demonstrate the repository functionality and the adopted meta-data schema, but now also to demonstrate its integration with the epidemic modelling and visualization software under development in other work-packages of the Epiwork project. In particular, the input files of the GleamViz simulator have been uploaded and their meta-data was characterized for inclusion in the Epidemic Marketplace. These files contain epidemic data about the H1N1 early phase, including:

- A dataset with the arrival times in different countries of the disease;
- Cumulative cases in Australia as of June 2009;
- Datasets with the early H1N1 outbreak timelines in 21 countries
- Two datasets with worldwide coverage (one including Mexico and another not including it, respectively);
- Datasets describing the fraction of imported cases for eight countries:

The repository also includes datasets gathered by the MEDCollector [11][27][28],

including sets of messages from the Twitter [25] website containing references to specific diseases and locations. These include the following datasets:

- Datasets of messages making reference to H1N1 influenza (from the 28th of May 2009 to the 22nd of March 2010) and to cholera (from the 5th of May 2009 to the 22nd of March 2010) for all the locations.
- Datasets for H1N1 influenza in four countries (Holland, Italy, Portugal and Spain).
- Other Twitter Messages containing the name of a disease and posted at a location.
- Email Messages containing the name of a disease and a location (including messages from ProMED-mail [12]).
- Messages from several CDC RSS feeds [18].
    i) CDC H1N1 Flu Updates
    ii) CDC Online Newroom
    iii) CDC E. coli Outbreak Updates
    iv) CDC Emergency Preparedness and Response: Recent Outbreaks and Incidents
    v) CDC Flu updates
    vi) CDC Salmonella: Outbreak updates
    vii) Emerging Infectious Diseases Journal
    viii) CDC Text Messages (ref)
    ix) CDC Travel Notices (ref)
    x) Health Information for International Travel / The Yellow Book
- Influenza activity estimates from Google Flu Trends [6] for all currently tracked countries and their capitals, when available.

Many of the catalogued resources, such as descriptions of Institutions or web sites, contain only meta-data, describing those resources and their location.

User Interface

The user interface had also a major overhaul since the release of the first EM prototype to the consortium. The initial user interface adopted the styles and templates defined by the software modules that we selected for use in the EM. The new interface adds a uniform look and feel to the entire system and significantly improves the usability of the EM

website (see http://www.epimarketplace.net).

The home page of the EM has access to the following functions:

- User Registration. Visitors to the website provide a description of their profiles and request an account at the EM. The request is conveyed to the operators in the back-end for validation or organizational elements;

- The main components of the EM, namely: Data Sets (connects to the Repository), Forum, Web Services (Mediator services) and Data Collection (MedCollector).

- Announcements, information on how to contact the EM developers and about the EM describing its main functions and goals.

User Interface Design

At the time of release to the public, the Epidemic Marketplace has a new interface that replaces the out-the-box styles provided by each of the components that we incorporated into the platform. Despite planning a fully revamped new interface for the Epidemic Marketplace in the coming months, we wanted to provide minimal visual consistency at the time of the release of the EM to the public. That consistency is crucial for the user to relate the various platform modules and perceive the location and meaning of their function. This is a requirement for improving all communication, information and navigational tasks. In other words, we wanted to quickly create a more functional, coherent and pleasant design.

The graphic layout, as it is well known, has an enormous impact in a web site's usability. With a proper and coherent interface the platform becomes more intelligible and understandable. For that, while structuring the information architecture, it is also mandatory to take into consideration elements, such as the main colours; font types and their weight, size or colour; text format and the position and sizes of the structural elements along the page.

On the Epidemic Marketplace, a graphic identity was the key requirement, and it had to be created so as to give some more visual consistency. The graphic identity itself is composed by several different graphic and informational elements, chosen according to the platform's informational nature.

The study began with the search for the web site's main colour. As the Epiwork logo had to

be visible in all the modules, it was picked a gradient variation from its letter "o", to create the main color. Several variations were tested on screen until getting the final one.

We considered that the Epidemic Marketplace is an online epidemic data organization platform while thinking about the possible graphic elements to integrate. Firstly, two key words were chosen to describe the platform's nature, in order to create some graphic elements for all the pages, providing visual consistency and coherence. Those words were "data organization" and "epidemic". Departing from those expressions, a graphic network's scheme was created, based on circles of several sizes and lines connecting them. The network of circles and lines refer both to data visualization networks and typical illustrations of epidemics.

Another key element for obtaining visual consistency and organization is the type font. In this case, for the Epidemic Marketplace, the chosen type fonts were Helvetica Neue Bold Condensed (for titles and major links) and Helvetica Regular, Italic, Bold and Bold Italic (for the plain text), due to their high reading efficiency on screen and balanced character. Sizes were chosen according to the text's importance and hierarchy on the page. The colours were also applied considering the text hierarchy or categorization, to better capture users' attention. Some typographic elements, such as underscores, were added as well, to accentuate visual coherence.

The homepage was developed from scratch, and all the other pages derived from it, incorporating the same graphic language. This includes the same main colour (a derivation from the 100% red) and header; type fonts or glyphs and their colours, size and format; number, position and colour of the main menu bars; description statements and the Epiwork logo position.

In the following pages, we provide some of the most relevant screenshots from the current user interface of the Epidemic Marketplace.
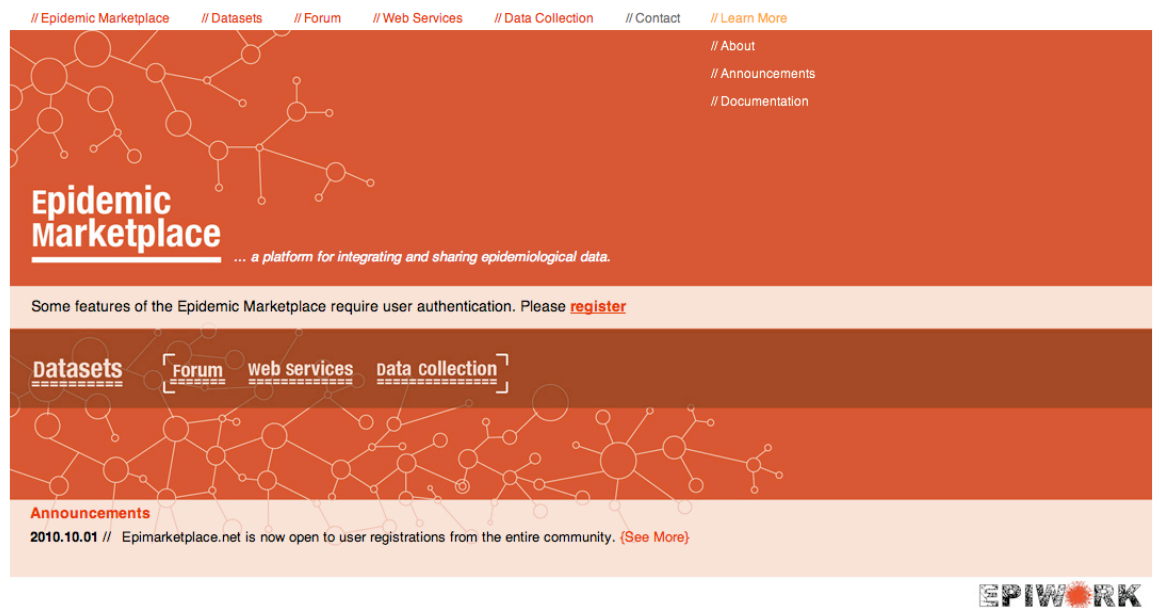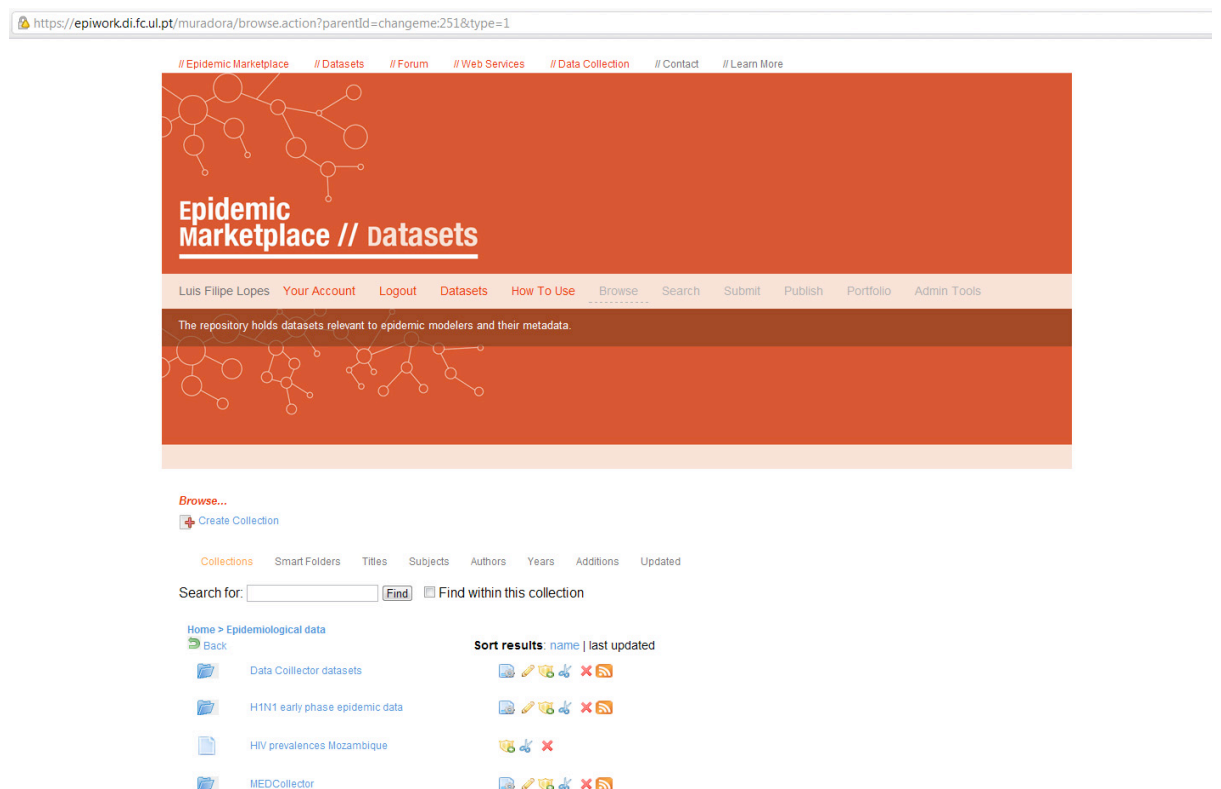
The Epidemic Marketplace Homepage:
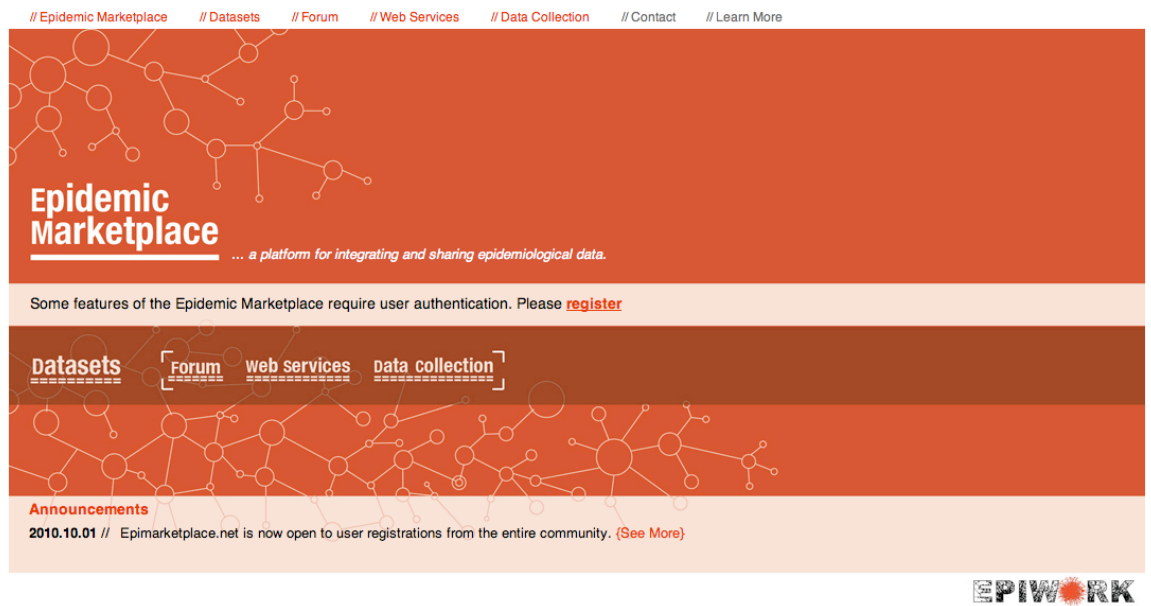
An example of the tooltips provided on mouse-over events on module links:
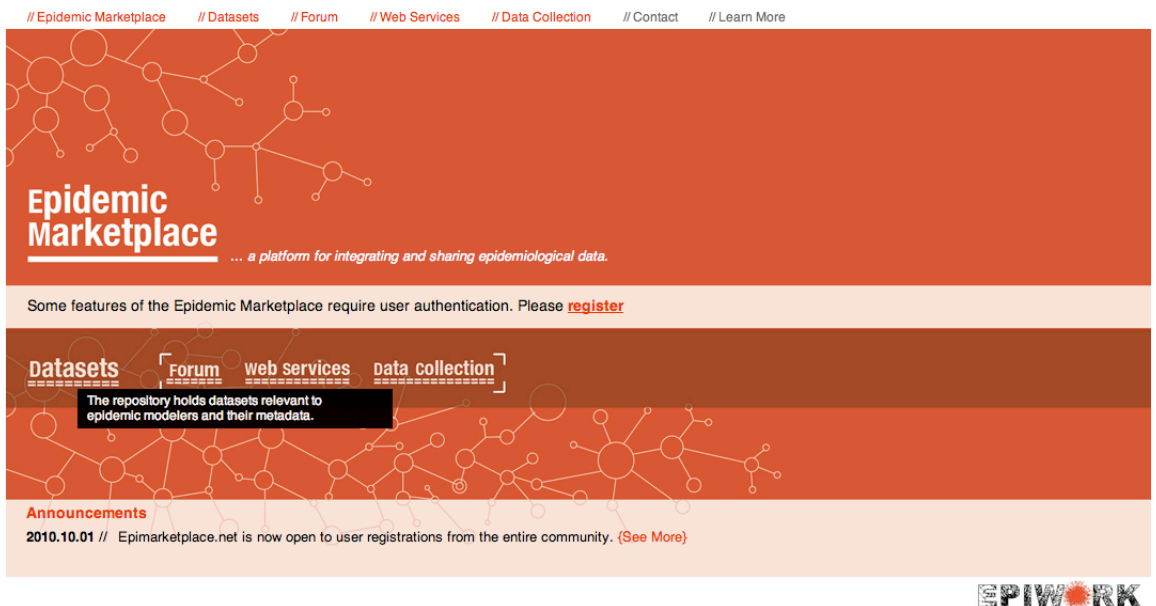
An Example of a Drop Down Menu:

Datasets (Digital Repository) page:

Forum Page:

Epidemic
Marketplace // Forum

Login   FAQ

The forum hosts discussions related to the available epidemic resources and services.

It is currently Thu Oct 28, 2010 11:52 am

Board index   All times are UTC

Forum	Topics	Posts	Last post
This board has no forums.			

Delete all board cookies I The team

Board index   All times are UTC

Who is online

In total there is 1 user online :: 0 registered, 0 hidden and 1 guest (based on users active over the past 5 minutes)

Web services Page:

Epidemic
Marketplace // web services

Search Objects   Search Collections   Fetch   Upload

The web services provide a programming interface for accessing epidemic resources.

EPIW☼RK

Data Collection Page:

Contact Information Page:

About Page:

// Epidemic Marketplace   // Datasets   // Forum   // Web Services   // Data Collection   // Contact   // Learn More

About Epidemic Marketplace //

The **Epidemic Marketplace** is a data integration platform where epidemiological data is stored, managed and made available, fostering collaboration.
It is developed under the Epiwork project, a multidisciplinary European research effort, aimed at developing the appropriate framework of tools and knowledge needed for the design of epidemic forecast infrastructures.

The **Epidemic Marketplace** includes four modules:

Datasets, a repository of user-submitted epidemic datasets with a search and navigation interface;
Forum, for discussions about epidemic resources;
Data Collection, for design and execution of workflows for harvesting epidemic data from multiple web sources;
Web Services, which provide an API for search, download and upload of epidemic datasets.

For further information, please visit the Epiwork web page at the University of Lisbon team.

Announcements Page:

// Epidemic Marketplace   // Datasets   // Forum   // Web Services   // Data Collection   // Contact   // Learn More

Announcements //

<2010.10.01>   Epimarketplace.net is now open to user registrations from the entire community.

<2010.03.15>   A new form is available which is closer to the Epidemic Marketplace metadata model, this form presents the user with some mechanisms to assist metadata creation.

<2010.01.05>   This first draft of the Epidemic Marketplace metadata scheme was presented for local discussion in the form of a DCAP. This scheme is being further analyzed and discussed in order to be implemented in the Epidemic Marketplace.

<2009.11.06>   New form update. This form includes fields specific to the EM metadata model.

<2009.10.11>   The user registration is now available.

<2009.05.24>   The Data Collector module is available. This module is capable of retrieving messages, containing epidemic related terms, from Twitter.

<2009.05.21>   We have just launched a first prototype version of a Repository for the epidemic marketplace. This repository was implemented with Fedora Commons 2.2.2 and Muradora 1.3.3.

Data Collection Documentation Page
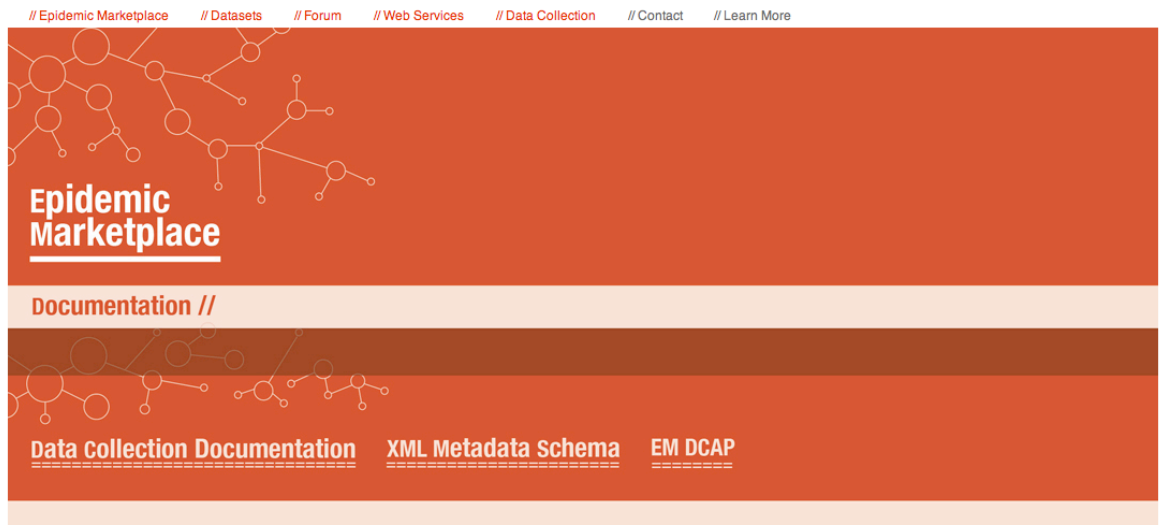
Conclusion

This report describes the work done in the period between the release to the consortium of the EM initial prototype and the public release of EM 1.0.

In addition to the online documentation at the EM website, the following publications describe the EM or its current configurations:

1. Mário J. Silva, Fabrício A.B. Silva, Luís Filipe Lopes, Francisco M Couto, Building a Digital Library for Epidemic Modelling. Proceedings of ICDL 2010 - The International Conference on Digital Libraries 1, p. 447–459, New Delhi, India, 23–27 February, 2010. TERI Press–New Delhi, India. Invited Paper.

2. Fabrício A.B. Silva, Mário J. Silva, Francisco M Couto, Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis. ERCIM News 82 – Special Theme: Computational Biology. July, 2010.

3. Luis Filipe Lopes, Fabrício A.B. Silva, Francisco M Couto, João Zamite, Hugo Ferreira, Carla Sousa, Mário J. Silva, Epidemic Marketplace: An Information Management System for Epidemiological Data. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.

4. João Zamite, Fabrício A.B. Silva, Francisco M Couto, Mário J. Silva, MEDCollector: Multisource Epidemic Data Collector. Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.

5. João Zamite, Multisource Epidemic Data Collector, Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.

6. Luis Filipe Lopes, A Metadata Model for the Annotation of Epidemiological Data, Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.

7. Hugo Ferreira, O Mediador do Epidemic Marketplace. Master Dissertation, University of Lisbon, Faculty of Sciences, September, 2010; (in Portuguese).

We have completed all the planned goals but one, the usage of the Drupal Content-Management System. As outlined in the Epiwork Deliverable D1.2, we also wanted to replace Muradora and the phpBB Forum by the Drupal Content-Management System [1][2]. as the new front-end for the EM repository. This task was actually harder to implement than anticipated, as it required upgrading Fedora Commons to a new major release and migrating all the content in the repository, in addition to adapting the new interface. We decided to stick with Muradora for the EM1.0 release. However, we already have a version of Drupal connected to the new version of Fedora Commons with the EM contents automatically migrated in our development environment.

In the next 6-9 months, we plan on releasing a new version 2.0 of the EM based on Drupal and a revamped user interface. In parallel, we are researching the implementation of a new access control scheme that would enable sharing of datasets under a social networking paradigm, and the negotiation of access rights to the datasets through the EM.

At the time of this report, the release of the EM 1.0 to the public was announced at the Epiwork Website (http://www.epiwork.eu). We plan to further disseminate the EM through further publications and announcements in web forums for epidemiologists.

References

[1] Byron A., Berry A., Haug N., Eaton J., Walker J. and Robbins J., Using Drupal, O'Reilly Media, December 2008. ISBN-0-596-51580-4

[2] Drupal website http://drupal.org/.

[3] Epidemic Marketplace at Google Code. http://code.google.com/p/epidemic-marketplace/..

[4] The GLEaM research team page . http://www.gleamviz.org/credits/.

[5] Google Analytics. http://www.google.com/analytics/

[6] Google Flu Trends. http://www.google.org/flutrends/.

[7] Lagoze C., Payette S., Shin E., Wilper C.. Fedora: an Architecture for Complex Objects and their Relationships. Inter. J on Digital Libraries Vol. 6, no. 2, pp 124-138, 2006.

[8] Lopes L.F. A Metadata Model for the Annotation of Epidemiological Data. Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010. http://xldb.di.fc.ul.pt/xldb/publications/Lopes:AMETADATAMODEL:2010_document.pdf

[9] Lopes L. F., Silva F. A. B., Couto F. M., Silva M. J., Epiwork D3.1: Meta-model Initial

Specification, Catalogue of Relevant Data, Platform Requirements. 30th September, 2009.

[10] Lopes L. F., Silva F. A. B., Couto F. M., Zamite J. M., Ferreira H., Sousa C., Silva M. J., Epidemic Marketplace: An Information Management System for Epidemiological Data. 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.

[11] Lopes L. F., Zamite J. M., Tavares B. C., Couto F. M., Silva F. A. B., and Silva, M. J. Automated Social Network Epidemic Data Collector. INForum informatics symposium. Lisboa, 2009.

[12] Madoff L. and Yu V. ProMED-mail an early warning system for emerging diseases. Clinical infectious diseases. Vol 39. No 2. Pp. 227-232. 2004.

[13] Nagios. http://www.nagios.org/

[14] Nguyen C., Dalziel J. Muradora: A Turnkey Fedora GUI Supporting Heterogeneous Metadata, Federated Identity, and Flexible Access Control. In: Proc. Third Intern. Conf. on Open Repositories, 2008.

[15] OAI, Open Archives Initiative . http://www.openarchives.org/.

[16] OASIS - eXtensible Access Control markup Language (XACML). http://www.oasis-open.org/committees/xacml/charter.php/..

[17] phpBB web site. http://www.phpbb.com/.

[18] RSS at CDC. http://www2c.cdc.gov/podcasts/rss.asp..

[19] Shibboleth web site. http://shibboleth.internet2.edu/.

[20] Silva F. A. B., Silva M. J., Couto F. M., Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis. ERCIM News 82 - Special Theme: Computational Biology. July, 2010.

[21] Silva M. J., Silva F. A. B., Lopes L. F., Couto F. M. Epiwork D3.2 Prototype of the Epidemic Marketplace Platform with an initial set of epidemiological databases integrated available to project participants. 30th September 2009.

[22] Silva M. J., Silva F. A. B., Lopes L. F., Couto F. M., Building a Digital Library for Epidemic Modelling. Proceedings of ICDL 2010 - The International Conference on Digital Libraries 1, p. 447–459, New Delhi, India, 23–27 February, 2010. TERI Press – New Delhi, India. Invited Paper.

[23] Takemura, C. and Crawford L. S., The Book of Xen, No Starch Press, 2009. ISBN-13 978-1-59327-186-2.

[24]   Tuttle S., Ehlenberger A., Gorthi R., Leiserson J., Owen N., Ranahandola S., Storrs M., Yang C. Understanding LDAP design and implementation. IBM International Technical Support Organization, 2nd ed., 2004.

[25]   Twitter. http://www.twitter.com/.

[26]   Xen. http://www.xen.org/)

[27]   Zamite J. M. Multisource Epidemic Data Collector. Master Dissertation, University of Lisbon, Faculty of Sciences, September 2010.
http://xldb.di.fc.ul.pt/xldb/publications/Zamite:MEDCollectorDissertation:2010_document.pdf.

[28]   Zamite J. M., Silva F. A. B., Couto F. M., Silva M. J., MEDCollector: Multisource Epidemic Data Collector.Proceedings of ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 August, 2010.