

The Multi-Knowledge Service-Oriented Architecture: Enabling Collaborative Research for E-health

Michele Amoretti

Francesco Zanichelli

Dip. di Ingegneria dell'Informazione, University of Parma, Italy

{michele.amoretti, francesco.zanichelli}@unipr.it

Abstract

The introduction of e-Health services is facilitating access to healthcare, regardless of geographical location or time, thanks to innovative telemedicine and personal health systems. E-Health is also breaking down many barriers, enabling health service providers (public authorities, hospitals) from different countries to work more closely together. In this context, the Multi-Knowledge platform may be regarded as a milestone, allowing geographically dispersed groups of biomedical researchers, dealing with different data sources as well as technological and organisational contexts, to create, exchange and manipulate new knowledge in a seamless fashion. Moreover, the Multi-Knowledge project defined a methodological framework for research experiments involving heterogeneous data, that can easily be extended to include additional sources of knowledge and expertise (biomedical data, images, environmental data), and can be applied to wider sectors of medical research.

1. Introduction

E-Health can be broadly defined as the use of Information and Communication Technologies tools and services to provide better healthcare. Whether e-Health tools are used behind the scenes by healthcare professionals, or even directly by patients, they play a significant role in improving the health of citizens. E-Health covers the interaction between patients and health-service providers, institution-to-institution transmission of data, or peer-to-peer communication between patients and/or health professionals. Examples include health information networks, electronic health records, telemedicine services, wearable and portable systems recording or communicating data, health portals, and many other ICT-based tools assisting disease prevention, diagnosis, treatment, health monitoring and lifestyle management.

The Multi-Knowledge project [1], which is funded by the European Commission in the context

of the Sixth Framework Programme for Research and Technological Development (thematic area Information Society Technologies), arises from the data processing needs of a network of medical research centres, located in Europe and USA, performing research activities in the field of metabolic and cardiovascular diseases. These needs are mostly related to the integration of three main sources of information, namely clinical data, patient-specific genomic/ proteomic data (in particular information acquired by means of microarray technology), and demographic data.

Critical and difficult issues addressed in the project concern the management of data which are heterogeneous in nature (continuous and categorical, with different order of magnitude, with different degree of precision, etc.), origin (statistical programs, manual introduction from an operator, etc.), and coming from different data environments (from the clinical setting to the molecular biology lab).

Still, the main objective of the Multi-Knowledge project is related to the development and validation of a collaborative IT platform for knowledge management, allowing geographically dispersed groups of researchers, dealing with different data sources as well as technological and organisational contexts, to create, exchange and manipulate new knowledge in a seamless fashion. Moreover, the Multi-Knowledge project defined a methodological framework for research experiments involving heterogeneous data, that can easily be extended to include additional sources of knowledge and expertise (biomedical data, images, environmental data), and can be applied to wider sectors of medical research.

The Multi-Knowledge platform enables workflow design and execution based on novel operating procedures to manage and combine heterogeneous data and make them easily available for data analysis. It is based on a service-oriented architecture, where functionality is grouped around business processes and packaged as interoperable services. The Multi-Knowledge platform is thus based on an IT infrastructure which allows different applications to exchange data with one another as they participate in

business processes. The aim is a loose coupling of services with operating systems, programming languages and other technologies which underlay applications.

The paper is organized as follows. In section 2 we discuss relevant literature on computer supported cooperative work (CSCW) and the role of service-oriented architectures in this context. In section 3 we recap the basic concepts of service-oriented architectures, which we deem necessary to fully appreciate the description of the MK platform provided in section 4, focusing on its functional properties which are made available by the synergetic interaction of several modules. In section 5 we describe the experiments which are being conducted across different sites. Finally, in section 6 the main results and contributions are summarized.

2. Related work

From the technical point of view, the MK platform is a Computer-Supported Cooperative Work (CSCW) system. Many researchers of information systems criticize current CSCW systems based on their usability and users' satisfaction. The Multi-Knowledge project has been entirely conducted with the supervision of end users, i.e. biomedical researchers, biostatisticians and also practicing physicians. This approach, suggested for example by Ruppel et al. [2], lead the consortium to obtain a higher-quality application, and better acceptance, which is particularly important in the case of collaborative systems.

The first feature demanded to the MK platform is knowledge sharing support. As suggested by Kindberg et al. [3], we distinguished between several types of knowledge: data, domain, users (their competences, their needs). The premise of MK-supported activities is that researchers from different organizations and institutes agree on sharing the (anonymized) data they have on their patients, and to exchange specialized knowledge. The issue of exchanging patient-related data is being partially solved by the increasing adoption of electronic medical records (EMRs), instead of traditional paper medical records (PMRs). Bringay et al. [4] observed that practitioners still prefer to use the PMR to collaborate, because electronic medical documents do not allow reproducing some practices of collaboration carried out with the PMR, in particular one practice: annotations which are used as support for the collaboration. In the research context of the Multi-Knowledge project, EMRs are required since clinical data must be automatically integrated with

genomic/proteomic data (in particular information acquired by means of microarray technology), and demographic data.

Knowledge sharing is just one kind of cooperative activity the MK platform was demanded to support. The other categories, according to Bardram's classification [5], are: organization of work, planning and scheduling, and communication, with the general objective of creating new knowledge. In the first phase of the project we explicitly detailed these activities, with particular emphasis on the identification of different roles for human actors (as we summarized in section 2). Role-Based Collaboration (RBC) theory is a natural approach to integrate the theory of roles into the CSCW systems [6,7,8].

From the specification of user requirements, the importance of the Process Modelling component (including Knowledge Extraction) of the Multi-Knowledge system definitely emerged. Workflow management technology is not used in healthcare as often as in other domains. Healthcare workflows have "transactional" elements, such as admitting a patient or taking a blood glucose measurement, but focusing on individual transactions obfuscates the most important element of healthcare workflows - the need to flexibly promote and maintain the highest possible standard of care for patients [9]. In the field of biomedical research, Workflow Management Systems (WMS) are seen as a viable solution for the creation and deployment of new flexible and extensible data integration and analysis network tools [10,11]. Some WMS have been proposed [12,13] and are now under careful testing aimed at the verification of their actual ability to cope with the data integration issue. While their potential is clear, some limitations are now arising, including both practical issues (e.g. quality of service, speed, access restrictions) and computational issues (e.g., long running jobs, huge input/output).

At the end of section 3 we discuss some service-oriented platforms for e-Health, compared with Multi-Knowledge objectives.

3. Service-oriented architectures

The Service-Oriented Architecture (SOA) [23] design style defines the use of loosely coupled software services to support the requirements of distributed applications. SOAs achieve loose coupling among interacting services by employing a small set of simple and ubiquitous interfaces to all participating software entities, together with

descriptive messages constrained by an extensible schema delivered through the interfaces.

In an SOA-based distributed environment, shared resources (applications and data, mainly) are made available on demand as independent services that can be accessed without knowledge of their underlying platform implementation. Service requestors can be end users (with client tools) or other services.

SOA promotes reuse at the macro (service) level rather than micro levels (e.g. objects), thus simplifying interconnection to and usage of existing IT (legacy) assets. This fundamental property derives from services being composable, stateless, discoverable and loosely bound together by a common compliance to a standard communications framework (while the encapsulated programming logic needs not to comply with any platform or technology set). In details, SOAs are message-oriented, since interoperability is achieved by means of messages, which are data formatted as XML documents and built compliant with the rules dictated by the contract associated with the service.

SOAs based on the traditional **client/server** paradigm are characterized by server applications, hosted by always-on end systems, which provide services to many other client applications hosted by sometimes-on end systems. Servers are often required to have static, well-known IP addresses.

In a **peer-to-peer** based SOA, arbitrary pairs of peer application entities communicate directly with each others. In a peer-to-peer architecture, none of the participant hosts is required to be always on; moreover, a participating host may change its IP address each time it comes on. Scalability is one of the greatest strengths of peer-to-peer architectures. For example, in a peer-to-peer file sharing application, most users are also downloading files that other users share, thus contributing vast storage and networking resources, so that the overall performance highly exceeds what can be offered by centralized servers. However, because of the highly distributed and decentralized nature of peer-to-peer applications, they can be difficult to manage. A service can be offered or withdrawn by a peer at any time. Clients have to discover whether, and where, a service is provided.

Several design patterns for SOAs have emerged, with reference to issues such as integration styles, messaging systems and system management. For a detailed description of these patterns, which is out of the scope of this paper, refer to [27].

3.1. Service workflow paradigms

Workflow is the operational aspect (process logic) of a work procedure: how tasks are structured, who performs them, what their relative order is, how they are synchronized, how information flows (routing rules) to support the tasks and how tasks are being tracked. Significant effort is being put into defining workflow patterns [24] that can be used to compare and contrast different workflow engines across both people-based and rule-based processes.

Current workflow solutions define client-server system architectures where the workflow server is monolithic and centralizes critical orchestration services such as process management, activity distribution, work list management, and directory services. Centralizing workflow functionalities offers important benefits (e.g. tracking), yet it becomes difficult to address the needs of distributed workflows on a WAN. Since the workflow reference model [25] introduced only the idea of distributed processes without defining a notation for expressing peer-to-peer interactions, later models have been proposed. The **orchestration model** provides a scope specifically focusing on the view of one participant, whereas a **choreography model** encompasses all parties and their associated interactions giving a global view of the system [26]. On one side, the choreographic model aims at constraining the behaviour of the services involved in the system ruling the exchange of their messages and distributing the state of the activity among the entities. On the other hand, orchestration is based on a central entity (the orchestrator) which carries out a business activity invoking other services and maintaining the global state.

3.2. Service-oriented platforms for e-Health

In the e-Health context, the COCOON [14] project is aimed at activating regional semantics-based healthcare information infrastructures with the goal of reducing medical errors. Another recent initiative is ARTEMIS [15], whose objective is to develop a semantic framework for the healthcare domain, building upon a peer-to-peer architecture in order to facilitate the discovery of healthcare services. Examples of such services are those listed by the Biological Web Services (BWS) page [16]. Among all, GeneCruiser [17] is a Web Service for the annotation of microarray data, developed at the Broad Institute (a research collaboration of MIT, Harvard and its affiliated hospitals). GeneCruiser allows users to annotate their genomic data by mapping microarray feature identifiers to gene

identifiers from databases, such as UniGene, while providing links to web resources, such as the UCSC Genome Browser. It relies on a regularly updated database that retrieves and indexes the mappings between microarray probes and genomic databases. Genes are identified using the Life Sciences Identifier standard. A more complex example of Web Service-oriented architecture providing transparent access to biomedical applications on distributed computational resources is the National Biomedical Computation Resource (NBCR) [18], which is based on Grid technologies such as Globus Toolkit. NBCR users are allowed to design and execute complex biomedical analysis pipelines or workflows of services.

Compared to Multi-Knowledge, all these initiatives lack in the definition of a collaborative platform in which different actors participate in the workflow execution, with experiment steps defined by and conducted under the responsibility of a principal investigator coordinating a research team which may include biomedical researchers, bioinformaticians, statisticians, etc.

4. The Multi-Knowledge platform

In this section we analyze the main functionalities of the Multi-Knowledge distributed platform (see figure 1), which is a service-oriented architecture (SOA) based on the peer-to-peer paradigm, in which services are provided by different nodes and (by default) are not listed in a centralized directory. The modularity of the system allows for a wide range of compositions of service-providing nodes, which may share their services by means of distributed publish/discovery mechanisms, but also using a centralized repository if the specific application does not present scalability issues.

Actually, the Multi-Knowledge platform integrates the following reusable modules:

- **MK-PORTAL**: platform entry point, report browser, etc.
- **MK-SEC**: security management tool
- **MK-DCNS**: data entry mask designer, data entry and integration tool
- **MK-DA**: statistical analysis and data mining tool
- **MK-VIZ**: visualization tool
- **MK-REP**: reporting tool
- **MK-WF**: workflow designer, workflow execution engine

We discuss how these modules are involved in knowledge management, data collection, data

analysis and workflow design and execution. We also provide some technical insights, even though basically all modules adopt Web Services in order to be interoperable and to facilitate the workflow engine in their orchestration.

4.1. Knowledge management

The Multi-Knowledge platform can be configured to cope with the usability needs of completely different users: practicing physicians involved in clinical data collection, biomedical researchers charged with genomic data normalization, principal investigators and biomedical researchers interested in browsing and downloading workflow descriptions and experiment reports. The entry point to the system is always the MK-PORTAL, which allows to obtain the MK-DA (which includes the MK-REP), the MK-WF, and the MK-VIZ, and to browse documents, reports, workflow templates, etc.

The MK-PORTAL module is based on two technologies: Joomla!, which is a popular Content Management System (CMS) and Web Application Framework (evolution of Mambo), e and Java Server Pages (JSP), which enables rapid development of web-based applications that are server- and platform-independent.

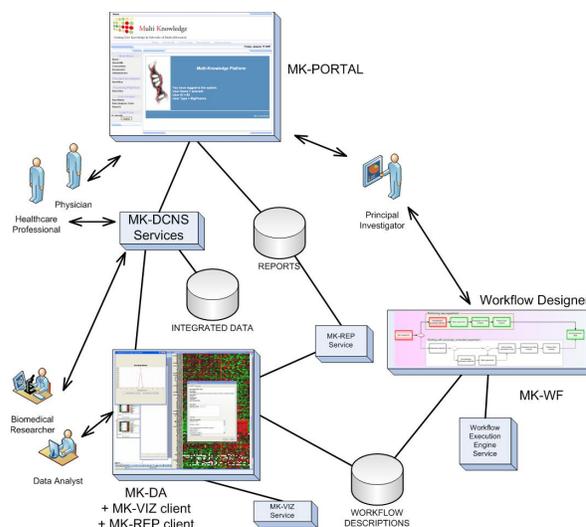


Figure 1. The Multi-Knowledge platform and the users which participate in the knowledge extraction process.

The MK-SEC module, developed in Java, is the glue between the MK-PORTAL and the other MK modules, which have been developed with J2EE/.NET technologies. The integration of Joomla!

and J2EE/.NET technologies is unusual, to the best of our knowledge. From this point of view, the MK-SEC is a pioneering artefact, defining a strategy for session sharing among Joomla!'s PHP pages and the JSP/ASP pages of the other MK modules, which are deployed in different web application engines. By achieving session sharing, it has been possible to introduce Single Sign-On (SSO), i.e. a method of access control that enables a user to authenticate once and gain access to differently deployed Multi-Knowledge modules. Another feature is role-based access, for which e.g. users with principal investigator access level are allowed to access most Multi-Knowledge modules, but they are not allowed to associate patients' clinical data with their vital statistics, for which only practicing physicians are responsible.

4.2. Data collection and normalization

The Data Collection and Normalization System (MK-DCNS) is composed by two sub-modules, namely Data Entry System (MK-DES) and Data Integration System (MK-DIS).

The data entry procedure, allowed by the MK-DES module, is straightforward (users in this context are practicing physicians). By clicking on the Data Entry link of the MK-PORTAL, a new window of the browser opens, showing the main page of the MK-DES. Here the user can perform a search (by first name, last name, or code) or create a new entry. The MK-DES forms accept clinical data (conforming to HL7 standards), as well as genomic and proteomic data resulting from microarray measurements. These are given as a set of feature extraction (FE) files, each one representing an experiment and containing all the data derived from the related microarray. Each expression FE file contains data on about 40000 genes and each protein FE file contains data on about 100 proteins. Typically, the first transformation applied to expression data, referred to as normalization, adjusts the individual hybridization intensities to balance them appropriately so that meaningful biological comparisons can be made. There are a number of reasons why data must be normalized, including unequal quantities of starting RNA, differences in labeling or detection efficiencies between the fluorescent dyes used, and systematic biases in the measured expression levels [28]. Furthermore, metabolomics data are given as tab delimited text files with two columns. The first column contains the metabolite description and the second column contains the corresponding numerical values and units. Finally, IMT and FMD data from each patient are entered to the system through the

MK-DCNS GUI. The MK-DIS module allows to generate and normalize a data matrix integrating clinical data, FE files, IMT and FMD data. Both MK-DES and MK-DIS rely on a set of Web Services running in a JBoss container.

4.3. Data analysis, visualization, and reporting

A Multi-Knowledge experiment is a set of successive data analysis cycles, aimed at extracting new knowledge from integrated heterogeneous data (clinical, demographical, genomic and proteomic), managed by a diverse set of researchers. Data analysis steps form the core of the experiment's analysis cycle. Through them, the data sample is successively analysed by different classes of researchers (having different "scientific cultures" and backgrounds) that use different analysis tools, work in different environments, at geographically dispersed sites.

The Data Analysis Tool (MK-DA) is an Eclipse plug-in, developed in Java, supporting several data mining processes, such as GO (Gene Ontology) analysis, classification, clustering, class discovery, and sequence motifs finding. The MK-DA is integrated with the Visualization Tool (MK-VIZ) module, which includes a Java-based Graph Generator component and a set of .NET-based Web Services that bring information related to the genes by accessing external genomic databases. The Graph Generator component is based on Mayday [19], an open source, Java-based tool for processing matrix related data and providing interactive data visualization. The Mayday platform has been fully customized in order to adapt to Multi-Knowledge requirements, input data set and desired output schemas and graphical plots and figures.

The MK-DA is also integrated with the Report Generator and Manager (MK-REP) module, including a Report Generator which builds PDF reports using the experiment description produced by the MK-WF module, together with data analysis results and images produced by the MK-VIZ tool. The MK-REP module also includes a ReportManager Web Service, which can be accessed by the MK-DA tool to store or retrieve reports in a specific database.

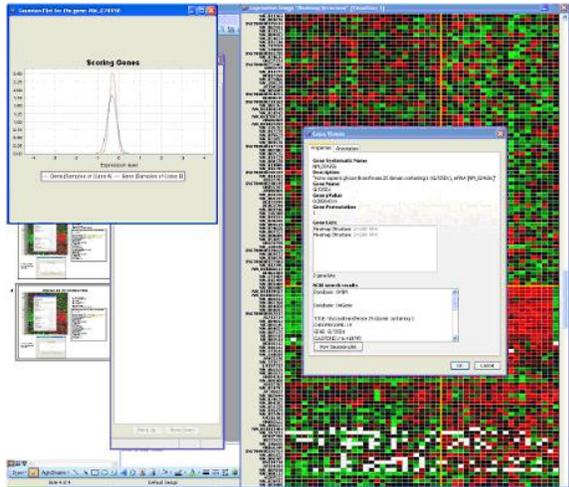


Fig. 2. MK-VIZ ranked heat map and gene information screens for gene NM_024656.

4.4. Workflow management

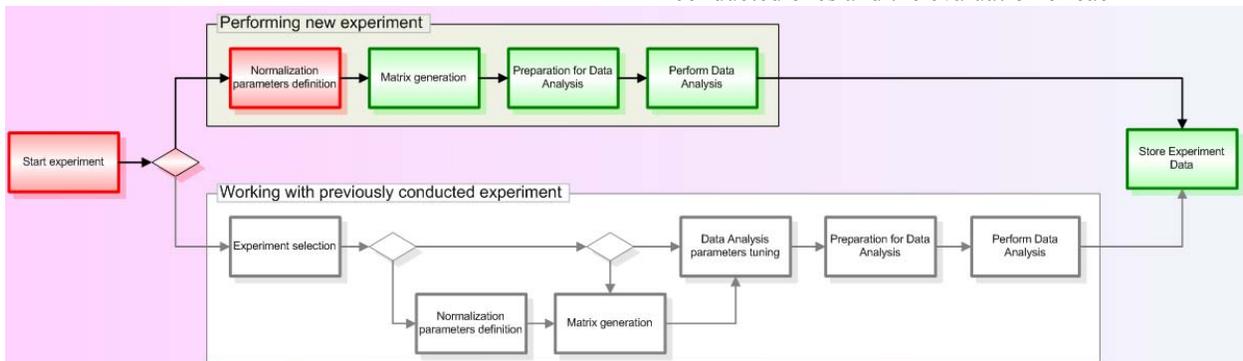


Fig. 3. Example of workflow design.

experiment is very important for the researcher to assess the value of each experiment and in the overall scientific process.

It is important to note again that the crucial objective of the Multi-Knowledge Process Modelling component is to guarantee the seamless integration of the different contribution brought in by the different research team members. This is not an always easy task because, for instance, some data analysis steps may be proposed/requested by researchers that do not have the specific expertise to conduct them while, vice versa, those who have the expertise to conduct them may not be able to understand the full, specific implications of the resulting knowledge.

The Workflow Designer and Execution Engine (MK-WF) allows principal investigators to define experiment steps to be conducted asynchronously or

We have illustrated how the MK platform supports data collection and normalization, as well as data analysis, visualization and reporting. In the following we focus on the orchestration of a whole experiment. As we already stated, a Multi-Knowledge research experiment consists of a set of experiment steps, defined by and conducted under the responsibility of a research team, coordinated by a Principal Investigator, that aims at achieving an predefined scientific goal.

Each data analysis step may generate new knowledge elements that contribute to create and successively expand an *experiment-related body of knowledge (EBoK)*. Based on an analysis of the EBoK (performed from their different scientific point of views) research team members can propose the execution of additional experiment steps or to further carry on the process. This means that we may see the process as a spiral in which every cycle allows a better focus on the scientific objective which was initially stated for the experiment. The combination of conducting experiments based on previously conducted ones and the evaluation of each

according to declared workflow patterns, passing control back and forth from different researchers. The experiments consist of dynamical cycles of data collection and analysis that aim at progressively achieving the scientific goal initially stated for the experiment.

When a team member, possibly after receiving a suggestion sent by another team member or by the principal investigator, decides to execute an experiment step, he/she:

- revises the proposed experiment step definition (a XML file) and possibly improves it based on her/his specific knowledge;
- executes the experiment step;
- generates a report, presenting the motivations for the experiment step as well as comments on step's execution and outcome.

During the execution of the experiment step, the workflow engine service activates the required modules (e.g. MK-DCNS for the data matrix generation, MK-DA for the data analysis, MK-VIZ for imaging, MK-REP for reporting). More specifically, the Workflow Engine service communicates with the appropriate modules of the platform in each experiment step, providing the necessary functionality to the user for accomplishing each specific step.

The MK-WF module interacts with the Data Collection and Normalization (MK-DCNS) module for setting the various experiment parameters for normalized data matrix to be generated. The MK-DCNS module generates the normalized matrix and stores it in a database.

Finally, the MK-WF module interacts indirectly with the Data Analysis (MK-DA) tool in the following cases:

- Retrieving the normalized matrix in order to be accessible from the MK-DA module running on the client machine.
- Checking if the MK-DA application is installed in the client machine. If it is not installed then the module downloads it from the MK-PORTAL and installs it on the client machine.

After data analysis has been performed and a report has been generated, the MK-WF module uploads the XML file containing the executed experiment log to the MK-PORTAL and stores the contained information to the Workflow database for future reference to the performed experiment.

Another functionality of the MK-WF module is that it enables users to conduct experiments based on previously conducted ones. In other words, previously conducted experiments can be loaded and new experiments can be executed after changing same parameters.

The MK-WF module was implemented under the Microsoft .NET Framework version 2.0. This includes the Workflow Execution Engine, implemented as a Web Service, as well as the workflow and knowledge extraction user interfaces, constructed as a Microsoft smart-client application. This approach combines the advantages of web applications with the enriched user interface of a native application and is suitable for user needs concerning experiments execution.

5. Pilot experiments

In a first instance of Multi-Knowledge pilot study, clinical, laboratory, instrumental and genomic

information has been collected from 50 subjects by the Department of Internal Medicine of the University of Parma. The sample has been used to validate the first Multi-Knowledge platform prototype, in particular the system modules related to data collection and normalization. Presenting biomedical results is out of the scope of this paper, but the interested reader can refer e.g. to [20].

The second instance of the pilot experiment has required further recruitment, up to a sample of about 150-200 subjects. This study is being performed to test a full-featured Multi-Knowledge system prototype, involving Internal Medicine Department at University of Parma, King's College London, Stanford University Medical Center, Technion Tel Aviv, University of Milan, and in conjunction with two related projects, namely the European Pocemon [21] and the Italian Sympar [22]. These partners are testing in particular the MK-WF and the MK-DA to exchange information about analysis workflow and to perform incremental data analysis as the pilot data set is modified.

6. Conclusions

The Multi-Knowledge platform is a relevant advancement in the e-Health landscape, supporting geographically dispersed groups of researchers to create, exchange and manipulate new knowledge in a seamless way. In the first part of the paper we presented relevant CSCW literature, focusing on knowledge sharing and workflow management systems. We compared the Multi-Knowledge project to other e-Health initiatives funded by the EU, which have similar purposes such as creating and sharing integrated biomedical information for better health. In the second part of the paper, we described the functionalities provided by the MK platform, with particular emphasis on services. Finally we summarized the activities carried out in the first pilot experiment, and those that are being conducted in the second (more complex) pilot experiment.

7. Acknowledgments

This work has been partially supported by the European Commission in the frame of the Project MULTI-KNOWLEDGE (Sixth Framework Programme, IST Priority, Grant #FP6-IST-2004-027106).

8. References

- [1] MULTI-KNOWLEDGE Consortium: project homepage. <http://www.multiknowledge.eu>
- [2] Ruppel, C., Konecny, J.: The role of IS Personnel in Web-based Systems Development: The Case of Health Care Organization. In: ACM SIGCPR Conference on Computer Personnel Research, pp. 130--135. ACM Press, New York (2000)
- [3] Kindberg, T., Bryann-Kinns, N., Makwana, R.: Supporting the shared care of diabetic patients. In: International ACM SIGGROUP Conference on Supporting Group Work, pp. 91--100. ACM Press, New York (1999)
- [4] Bringay, S., Barry, C., Charlet, J.: Annotations: A Functionality to support Cooperation, Coordination and Awareness in the Electronic Medical Record. In: Cooperative Systems Design (COOP '06), pp. 39--54. IOS Press, Amsterdam (2006)
- [5] Bardram, J. E.: Collaboration, Coordination, and Computer Support, An Activity Theoretical Approach to the Design of Computer Supported Cooperative Work. PhD Thesis, University of Aarhus, Denmark (1998)
- [6] Edwards, W.K.: Policies and Roles in Collaborative Applications. In: ACM Conference on Computer-Supported Cooperative Work (CSCW'96), pp. 11--20. Cambridge, USA (1996)
- [7] Guzdial, M., Rick, J., and Kerimbaev, B.: Recognizing and Supporting Roles in CSCW. In: ACM Conference on Computer-Supported Cooperative Work (CSCW'00), pp. 261--268. Philadelphia, Pennsylvania, USA (2000)
- [8] Smith, R. B., Hixon, R. and Horan, B.: Supporting Flexible Roles in a Shared Space. In: ACM Conference on Computer-Supported Cooperative Work (CSCW'98), pp. 197--206. Seattle, Washington, USA (1998)
- [9] LeMoine, D.: Going with the Flow: Interaction Design for Healthcare. In: Journal of Design, Cooper Consulting (2003)
- [10] Stein, L.: Creating a bioinformatics nation. In: Nature, No. 417, pp. 119--120 (2002)
- [11] Romano, P., Rasi, C. and Marra, D.: The automation of bioinformatics processes through workflow management systems. In: Seventh Spanish Symposium on Bioinformatics and Computational Biology (JdB06), Zaragoza, Spain (2006)
- [12] Oinn, T., Addis, M. et al.: Taverna: a tool for the composition and enactment of bioinformatics workflows. In: Bioinformatics, Vol. 20, No. 17, pp. 3045--3054 (2004)
- [13] Stevens, R., Robinson, A. and Goble, C.: myGrid: personalised bioinformatics on the information grid. In: Bioinformatics, Vol. 19, No.1, pp. 302--304 (2003)
- [14] COCOON consortium: project homepage. <http://www.cocoon-health.com>
- [15] ARTEMIS consortium: project homepage. <http://www.srdc.metu.edu.tr/webpage/projects/artemis/index.html>
- [16] Hull, D.: The Biological Web Services page. <http://taverna.sourceforge.net/index.php?doc=services.html>
- [17] Liefeld, T., Reich, M., Gould, J., Zhang, P., Tamayo, P. and Mesirov, J. P.: GeneCruiser: a Web Service for the annotation of microarray data. In: Bioinformatics, Vol. 21, No. 18, pp. 3681--3682 (2005)
- [18] Krishnan, S., Baldrige, K., Greenberg, J., Stearn, B. and Bhatia, K.: An End-to-end Web Services-based Infrastructure for Biomedical Applications. In: 6th IEEE/ACM International Workshop on Grid Computing, Seattle, Washington, USA (2005)
- [19] Dietzsch, J., Gehlenborg, N., and Nieselt, K.: Mayday - a microarray data analysis workbench. In: Bioinformatics 2006 22(8):1010-1012.
- [20] Steinfeld, I., Navon, R., Ardigò, D., Zavaroni, I. and Yakhini, Z.: Semi-supervised class discovery using quantitative phenotypes – CVD as a case study. In: BMC Bioinformatics, 8(Suppl 8):S6 (2007)
- [21] Pocemon project description at PCS' homepage. <http://www.pcs.at/index.php?id=38&L=1#c193>
- [22] Sympar Consortium: project homepage. <http://www.sympar.org>
- [23] He, H.: "What is Service-Oriented Architecture", <http://webservices.xml.com/pub/a/ws/2003/09/30/soa.html>, September 2003.
- [24] van der Aalst, W.M.P., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow Patterns. In: Distributed and Parallel Databases, 14(3), pp. 5--51, July 2003.
- [25] Hollingsworth, D.: The Workflow Reference Model. Workflow Management Coalition, January 1995.
- [26] Peltz, C. : Web services orchestration and choreography, IEEE Computer, Volume: 36, Issue: 10, pp 46- 52, October 2003.
- [27] Erl, T.: SOA Design Patterns. Prentice Hall, 2008.
- [28] Quackenbush, J.: Microarray data normalization and transformation. Nature Genetics Supplement, 32, pp. 496--501, December 2002.