GATiB-CSCW, Medical Research Supported by a Service-Oriented Collaborative System

Konrad Stark[1], Jonas Schulte[2], Thorsten Hampel[2], Erich Schikuta[1], Kurt Zatloukal[4], and Johann Eder[1,3]

[1] University of Vienna, Dept. of Knowledge and Business Engineering, Austria
[2] University of Paderborn, Heinz Nixdorf Institute, Germany
[3] University of Klagenfurt, Dept. of Informatics Systems, Austria
[4] Medical University Graz, Institute of Pathology, Austria

Abstract. Medical research is a collaborative process in an interdisciplinary environment that may be effectively supported by a Computer Supported Cooperative Work (CSCW) system. Such a system imposes specific requirements in order to allow flexible integration of data, analysis services and communication mechanisms. Persons with different expertise and access rights cooperate in mutually influencing contexts (e.g. clinical studies, research cooperations). Thus, appropriate virtual environments are needed to facilitate context-aware communication, deployment of biomedical tools as well as data and knowledge sharing. We systematically elaborate the main requirements of a medical CSCW system and present a conceptual model, as well as an architectural proposal satisfying the demands. We design a prototypical virtual workbench to support research and routine activities in the context of the GATiB (Genome Austria Tissue Bank) initiative.

Keywords: Medical Research, CSCW, Service-Oriented Architecture.

1 Introduction

Medical care and medical research are cross-fertilising areas. Courses of disease may be monitored and analysed from a scientific perspective, whereas treatment development and medication design benefit from the the results of research. Vast amounts of patient records containing information about diagnosis, laboratory tests, radiology images and medications are created continuously. For example, a medium-sized hospital in Austria provides medical treatment for 512,000 patients within one year (*http://www.klinikum-graz.at*). Although most of the data is recorded in medical information systems, they do not support collaborative work. Thus, there is a strong need to share, contextualise and annotate data allowing inter-organisational and interdisciplinary collaboration. A wide range of biomedical applications have been developed for very specific purposes, e.g. image processing, gene expression analysis. However, they also lack support for collaborative processes. In this paper we designed a *Computer Supported Cooperative Work (CSCW)* system that meets the requirements of the medical research

Z. Bellahsène and M. Léonard (Eds.): CAiSE 2008, LNCS 5074, pp. 148–162, 2008.
© Springer-Verlag Berlin Heidelberg 2008

domain, especially out of the field of flexible service integration (SOA) and object oriented data integration (integration of persistency layers). Standard “out of the box“ CSCW systems do not meet these requirements. Therefore, some new architectural elements had to be developed. Collaboration in the medical research field is characterised by high complexity and high variation of the collaborative situations. Data is distributed over several institutes and underlays various restrictions in accessibility for different persons (roles). Access to patient data is for example highly limited by the context or role of application. Data is collected, restructured, analysed and shared with each other in different settings. This paper addresses such questions by defining the requirements for a CSCW system. The resultant architecture of our CSCW system is also presented to highlight how the requirements can be attained in the field of medical research. However, in its characteristic as a collaborative knowledge management system our GATiB-CSCW system can be easily used as an e-learning platform. Our means of flexible configuration of access rights and the architectures ability to define contextual views on the presented data allows it to be used as a powerful collaborative learning platform. Much work has been done in developing IT infrastructure for the biomedical research. Some grid-based platforms support effective co-working of researchers on distributed data sets [1], expose various data sources on the grid and allow access by web services [2] or enhance interoperability between distributed medical models in a grid portal [5]. Other approaches provide an infrastructure for collaboratively using distributed computational services and data resources [8,6,12]. Though, few approaches [9,4] exist to explicitly support collaboration processes in the biomedical context, we encourage to close the gap by implementing a collaboration-aware biomedical infrastructure based on a service-oriented architecture. This work was developed in the context of the biobank initiative GATiB (Genome Austria Tissue Bank) which is part of the Austrian Genome Program (*http://www.gen-au.at*). GATiB aims at the establishment of a tissue bank which builds on a collection of diseased and corresponding normal tissues representing all diseases at the natural frequency of occurrence from a non-selected central European population of more than 700.000 patients. Major emphasis is placed on annotation of archival tissue with comprehensive clinical data including follow-up. A more detailed description of the biobank initiative is given in [3]. The paper is structured as follows: To describe the background of the requirements for a CSCW system useable for medical research, Section 2 presents example collaboration scenarios and summarises the main requirements for such a system. Some practical examples and a prototypical virtual workbench is presented in Section 3. The conceptual components for the GATiB-CSCW system are described in Section 4. The architecture of the CSCW system is specified in Section 5. Our paper ends with a description of the current status and an overview of ongoing and future work in Section 6.

2 Scenarios and Requirements

Collaboration in the GATiB project focuses on both the medical research and routine activities of medical scientists and supporting staff. Data that is locally distributed over institutes and research groups is accessed in manifold ways and for various purposes. The following three example scenarios describe different types of collaboration in medical research.

Scenario 1: A patient diagnosed with mamma carcinoma is assumed to suffer from a rare subtype of breast cancer. A group of expert pathologists are concerned with the correct cancer classification. Each of the experts needs detailed access to anamnesis data of the patient and her family. Additionally, virtual slides made from the instantaneous section are used to cooperatively mark and annotate sections of the image. A set of similar cases as well as selected state-of-the-art publications are provided for comparative analysis.

Scenario 2: For an extensive evaluation of the course of disease of liver cancer over the past 20 years information about diagnoses, family anamnesis and follow-up data (medication, therapy, resection) of relevant cases is needed. Data from the institutes pathology, oncology and surgery is accessed as well as survival data in order to support statistical analyses. The distribution of liver cancer subtypes is calculated and an inventory of the biobank lists all associated tissue samples. The results are used both in revision reports as well as in a publication.

Scenario 3: Due to a cooperation between the hospital and a pharmaceutical company a group of suitable human-tissue donors is to be identified to support a drug discovery study. Therefore, pathological diagnoses, survival data and tissue images of patients that have signed an informed consent are required. After searching and structuring the information confidential patient data has to be protected since an external organisation is involved in the study. Hence, identifying attributes (name, day of birth) are eliminated and quasi-identifying attributes are k-anonymised [10,14]. Life style data is included by filling out questionnaires. Further, a tissue microarray of the relevant cases is made in order to test candidate tumour markers. The results should be made available to other research groups.

These scenarios demonstrate the diversity of collaboration types that may occur. A cooperative system in a biomedical research environment requires a high degree of flexibility and extensibility. Distributed data is accessed in different levels of grading considering data privacy issues, it is annotated and analysed. We use the above-mentioned scenarios to deduce general requirements a CSCW system has to comply with.

Requirements

- $R(1)$ **User and role management.** The CSCW has to be able to cope with the organisational structure of the institutes and research groups of the hospital. Data protection directives do have to fit in the access right model of the system. Though, the model has to be flexible to allow the creation of new research teams and information sharing across organisational borders.
- $R(2)$ **Transparency of physical storage.** Although data may be stored in distributed locations, data retrieval and data storage should be solely

dependant on access rights, irrespective of the physical location. That is, the complexity of data structures is hidden from the end user. The CSCW system has to offer appropriate search, join and transformation mechanisms.

- $R(3)$ **Flexible data presentation.** Since data is accessed by persons having different scientific background (biological, medical, technical expertise) in order to support a variety of research and routine activities, flexible capabilities to contextualise data are required. Collaborative groups should be able to create on-demand views and perspectives, annotate and change data in their contexts without interfering with other contexts.
- $R(4)$ **Flexible integration and composition of services.** A multitude of data processing and data analysis tools exists in the biomedical context. Some tools act as complementary parts in a chain of processing steps. For example, to detect genes correlated with a disease, gene expression profiles are created by measuring and quantifying gene activities. The resulting gene expression ratios are normalised and candidate genes are preselected. Finally, significance analysis is applied to identify relevant genes [16]. Each function may be provided by a separate tool; for example by Genespring®and Genesis®[7,15]. In some cases, tools provide equal functionality and may be chosen as alternatives. Through flexible integration of tools as services with standardised input and output interfaces a dynamic composition of tools may be accomplished. From the systems perspective services are technology neutral, loosely coupled and support location transparency [11]. That is, the execution of services is not limited to proprietary operation systems and any service caller does not know the internal structure of a service. Further, services may be physically distributed over departments and institutes, e.g. image scanning and processing is executed in an own laboratory where the gene expression slides reside.
- $R(5)$ **Support of cooperative functions.** In order to support collaborative work suitable mechanisms have to be supplied. One of the main aspects is the common data annotation. Thus, data is augmented and shared within a group and new content is created cooperatively. Therefore, Web 2.0 technologies like wikis and blogs procure a flexible framework for facilitating intra- and inter-group activities.
- $R(6)$ **Data-coupled communication mechanisms.** Cooperative working is tightly coupled with excessive information exchange. Appropriate communication mechanisms are useful to coordinate project activities, organise meetings and enable topic-related discussions. On the one hand a seamless integration of email exchange, instant messaging and voip tools facilitates communication activities. We propose to reuse the organisational data defined in $R(1)$ within the communication tools. On the other hand, persons should be able to include data objects in their communication acts. E.g., images of diseased tissues may be diagnosed cooperatively, whereas marking and annotating of image sections supports the decision making process.
- $R(7)$ **Knowledge creation and knowledge processing.** Cooperative medical activities frequently comprise the creation of new knowledge. Data sources are linked with each other, similarities and differences are detected,

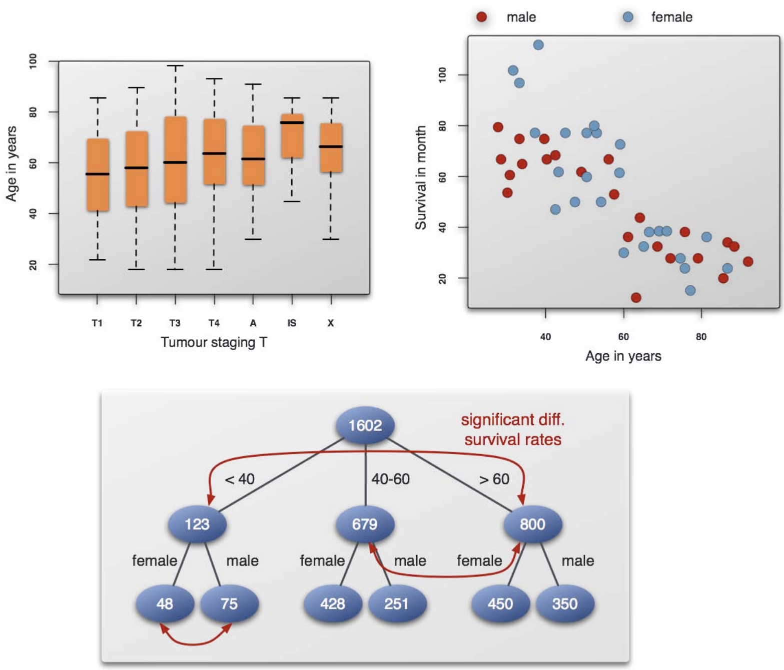

Fig. 1. TUMour Staging, Age, Survival Plots

and involved factors are identified. Consider a set of genes that is assumed to be strongly correlated with the genesis of a specific cancer subtype. If the hypothesis is verified the information may be reused in subsequent research. Thus, methods to formalise knowledge, share it in arbitrary contexts and deduce new knowledge are required.

3 Practical Examples and Virtual Workbench

In order to illustrate different types of data and knowledge we have to cope with in the GATiB project we give some practical examples. A key aspect of the biomedical collaboration is analysing the data collected about tumours in multiple ways. In Figure 1 three examples of graphical presentation of patient records are given. In the top left part of the figure the age distribution of patient groups is visualised by boxplots. Patients are grouped by tumour staging T (TNM classification). The top right part illustrates survival periods and age of female and male patients by a scatterplot. These visualisations may be a helpful presentation for scientists to formulate new hypotheses. The stratification tree at the bottom gives a detailed characterisation of the whole data set. This presentation is useful to survey frequencies of attribute value combinations. The original stratification tree was augmented with results of statistical tests. Hence, significant differences among subgroups of the data set are displayed by red arrows. In our example there is a significant difference of survival periods between male and female patients under 40. The dependency is obvious when looking at the scatterplot of Figure 1. Further, the survival period of patients older than 60 is significantly different from both younger age groups. Various types of data and knowledge presentation may serve different purposes. A tabular presentation of knowledge may be useful for final documentation, either to be exported into publication drafts or into presentations, while the visualisation by stratification trees may support generation of hypotheses in ongoing research. We propose to introduce formal representations of detected relationships to capture knowledge structures that may be reused in similar contexts and constitute the basis for reasoning in medical decision support systems. The details of knowledge representation are topics of ongoing research and beyond the scope of the paper. In Figure 2 a prototypical view of the scientific workbench is given. A group of persons cooperate in a virtual environment in the context of an annotation project. All project members currently working in the *context* of the project are marked as online. Online persons may communicate immediately by instant messaging. Additionally, further contact information (phone number or mail address) may be retrieved by clicking on the second icon. A set of records of liver cancer patients represents the shared data objects. Individual objects may be accessed, altered and marked as 'completely annotated'. A set of relevant functions is displayed to visualise and analyse data. Further, related publications are presented and discussion forums (data and knowledge topics) are available.

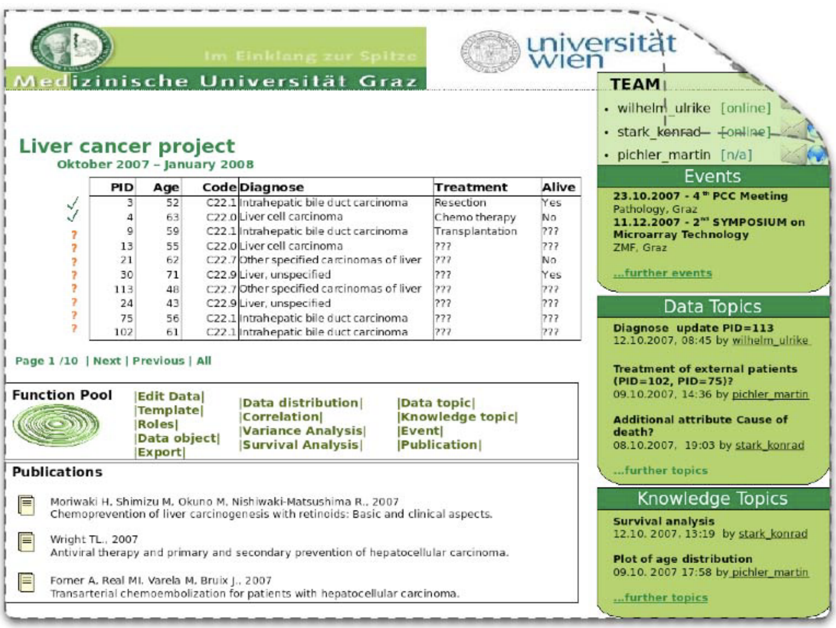

Fig. 2. Scientific Workbench

4 Conceptual Model

The example applications show that data used for collaboration has to be differentiated. Furthermore, the environment for a collaboration has to be specified and the users have to be represented in the CSCW system in a suitable way. We elaborated a conceptual model in which we specify how the requirements of the previous section could be met.

4.1 Resource

The example applications described in Section 2 highlight that the same data can be analysed and visualised in various ways. Our GATiB-CSCW system has to distinguish between the basic data on which the analyses are performed and the results of analyses. A resource is either a data or knowledge object that is accessed in a cooperative process. While data objects consist of restructured or transformed data, knowledge objects are built by deducing relationships and conclusions from data objects. Thus, a knowledge object always has at least one associated data object. Data and knowledge objects have some common characteristics: the content of both data and knowledge objects may be personal. Access to data objects can be restricted due to legislative regulations to guarantee data privacy. Further, the content of knowledge objects is to be protected in cases of ongoing research work. Moreover, both data and knowledge objects have to be protected from unauthorised modifications. The CSCW system needs a user management handling authentication and authorisation before accessing data. To protect the content of knowledge objects in ongoing research, a version management is required.

Data Object. We consider a data object as an information that was extracted from information systems or documents or was entered manually. In the context of the GATiB project emphasis is put on biomedical data. Biomedical data consists of data records from medical hospitals and research facilities that are of interest for collaborative research. It includes patient-specific records extracted from medical information systems or data manually entered as a result of an annotation process. Data may originate from clinical studies, anamnesis, lifestyle, and survival data as well as gene expression profiles. We do not put limits on the type of data nor on the size. Hence, a data object may consist of a set of database records that was returned as a result of a specific query linked with supplemental images as illustrated in Figure 3. Further, a data object may be a text document containing fragments of a publication. The CSCW system must not have limitations concerning the data stored and accessed. Since arbitrary data types have to be supported, applications used to visualise or modify the data should be made available within the CSCW system. We propose flexible levels of data granularity to satisfy requirement $R(2)$.

Knowledge Object. A knowledge object is the result of a non-empty sequence of functions applied to at least one data object. Generally, each knowledge object enhances the information that is currently available in data objects. We

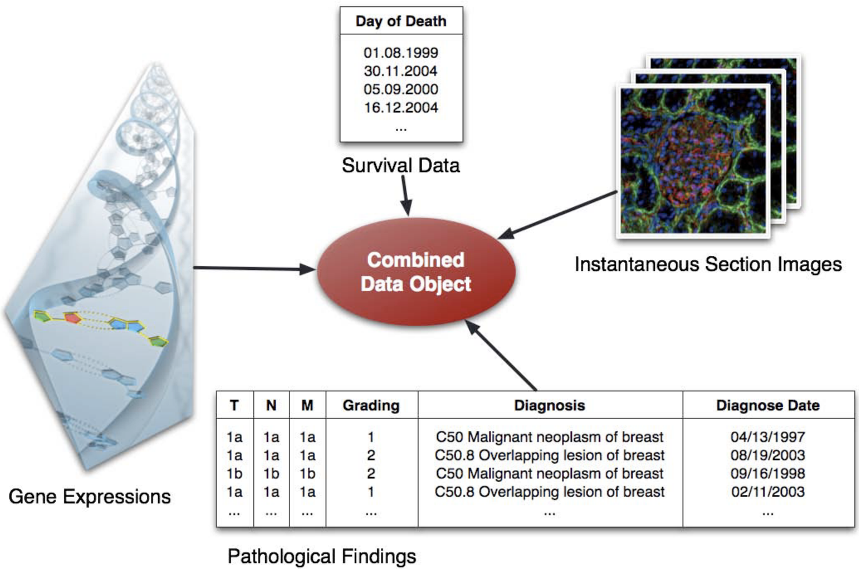

Fig. 3. Combined Data Object

distinguish various types of knowledge objects: graphical objects (plots, diagrams, etc.), tabular objects presenting summarised information (e.g. the results of aggregation operations) and knowledge structures (e.g. annotations). Semantic structures may be used to formalise knowledge. Relationships between objects or groups of objects are stored in a processible way in order to build a knowledge repository. The repository integrates conclusions from various contexts allowing the user to explore consolidated knowledge and deduce new knowledge. Making knowledge persistent and processable allows us to fulfil requirement $R(7)$. The post-processing of accessible data objects is a specific but also important criterion for a medical CSCW system research. Since there should be no limitations concerning the data stored and accessed, the CSCW system has to support all required post-processing steps and must be flexible to integrate new functions.

4.2 Collaboration Context

A collaboration context specifies the type of collaboration. It provides the general framework of relationships between individuals, data and knowledge. It may be considered as a basic template defining how data and knowledge is accessed and modified by individuals in a certain type of collaboration. In our collaborative medical research system we use three basic types of contexts: **patient-centred**, **project-centred** and **disease-centred** context. These patterns are applicable to the scenarios described in Section 2 which correspond to common use cases of medical routine work as well as medical research activities.

Generally, the patient-centred and project-centred contexts access more sensitive data. A patient-centred context starts with the patients data and will be used for medics working with the patient directly. Hence, involved individuals have to have access rights to original patient-related data. Whereas in the disease-related context access is given to anonymised and/or summarised data of the same disease type. As a consequence, instead of focusing on specific patients, a particular disease type, its related therapies and medications are of main interest for the collaboration act. The project-centred context provides a specific environment to coordinate collaborative actions (e.g. data sharing and annotation, discussion forums, communication tools etc.). The collaboration context conforms to an environment which present users a common background for their work. It presents data needed as background information for their collaboration since all data available in the context can be assumed as known from the users. It is obvious that a collaboration context must consider authentication and authorisation in order to limit the user's capabilities to access and modify data. Further important is the view on the data available in a collaboration context. Section 4.5 details the need for different views on the same data. The CSCW system has to be able to present these three context types. Though, to develop a flexible system, it has to support the definition and usage of arbitrary contexts.

4.3 Knowledge Spaces

Knowledge spaces are our representation and structure of presenting the different context types of collaboration situations to the users/actors in the collaboration process. Users need a knowledge space as a virtual environment for their collaboration activities in which they can meet and work. We define a knowledge space as an *actual use case* in a *collaboration context*. In scenario 3, a knowledge space is set up for those persons participating from the hospital and the pharmaceutical company, which are therefore a member of this specific collaboration group. Although knowledge spaces are separate virtual concepts, data and knowledge exchange between knowledge spaces is encouraged. An important capability is to upload data or link to remote data which is already available. To suffice the individual users' needs for their collaboration, they should organise and structure the knowledge space completely on their own. Consequently, users are responsible for the organisation of their knowledge space which implies highest flexibility. However, reorganisation of one specific knowledge space might be restricted to a limited group of users in order to avoid unauthorised modifications. Such self-organisation conforms to the self-organisation forms of knowledge in the Web 2.0 (tagging). Further concepts of the Web 2.0 like annotating available information are important for a useful CSCW system. Knowledge spaces are extendable in order to invite other users of the CSCW to join the collaboration.

Communication is the most important criterion to perform a successful collaboration: In particular, supporting discussions is crucial for the usability of a CSCW system since it is an important tool in the context of collaboration between users located at different places or contributing at different times. To offer a useable platform for collaborative work, different communication ways have to be supported. A classical and popular form of synchronous communication is to use instant messaging. Initially, the chat protocol is only privately accessible for the participants, however, they can make it accessible for all users of the knowledge space. In the context of a platform for medical research, discussion forums for knowledge and data topics are helpful. Furthermore, Web 2.0 concepts like wikis and blogs should be supported by the CSCW system. Wikis can be generated for a dedicated collaboration context or be accessible for a subset or all individuals of the CSCW system. To add comments to knowledge objects or data objects, the CSCW system offers annotation capabilities. Thereby, data and knowledge objects can be annotated in the same way and the CSCW system is responsible for granting access to annotations according to the users' access rights. Hence, we may fulfil requirements $R(5)$ and $R(6)$. Realising user-awareness in the CSCW system is crucial to support communication. User-awareness implies that users are represented in the virtual world by avatars, photos, or an icon. This enables other users to see which individuals are registered, are author of a wiki entry, annotation etc., are currently logged in, and can be contacted.

4.4 Individual

An individual is a person participating in collaborative acts. Individuals may be for instance researchers from the medical or biological domain, medical students or project managers. Individuals are categorised into internal and external persons in order to differentiate between access to sensitive patient-related and research-related data and access to anonymised and summarised data. Internal individuals are for example employees of the Medical University Graz and external individuals may be members of companies or research institutes cooperating with the university. The CSCW system has to ensure that individuals are distinguished in order to reflect the activities of one individual in the real world as a one-to-one mapping in the virtual world. This is important in the context of collaboration (discussions, annotations, ...) as well as for the data access (authentication and authorisation). The organisational mapping accomplishes the user and role management specified in $R(1)$.

4.5 Roles

Individuals having the same access rights and using the same set of functions may be integrated to groups and certain roles may be assigned to users/groups as illustrated in Figure 4. Roles are used to adjust the access of individuals and groups to resources. That is, a role is used to bundle access parameters that are classified into three different categories: **view parameters**, **modification parameters** and **extension parameters**. Flexible presentation of data and data annotation capabilities satisfy requirement $R(3)$. Individuals can be part of several groups and a group can consist of subgroups and roles may be assigned to both individuals and groups. Further, roles are coupled with knowledge spaces allowing to access a resource in various contexts.

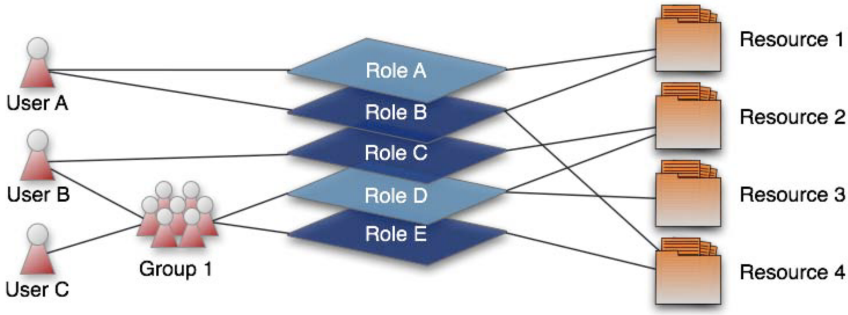

Fig. 4. Role Assignment

4.6 View Parameters

Different knowledge spaces have a different focus on the same data. This implies that the data is visualised differently according to the selected knowledge space. View parameters are utilised to specify how the data is presented. Although the same data is shared among a group of individuals, sometimes part of the data is to be hidden or presented in an alternative way. For instance, consider sensitive information like patient name and day of birth. In a clinical study the identifying attributes of a patient record have to be presented to assign further attributes like radiology parameters correctly. In order to guarantee data privacy, these attributes should not be presented to supporting staff like medical students. Hence, the identifying attributes are suppressed for supporting staff and displayed for scientific staff by setting the view properties in the corresponding roles appropriately. The CSCW system has to support a fine-grained definition of view parameters. Supporting the self-organisation of the knowledge spaces, view parameters must be able to be defined by the users themselves. Here, the definition might be available for a specific group of users that are responsible for the organisation of the knowledge space.

4.7 Modification Parameters

The content of a resource may be changed during a review, discussion, modification or annotation process. Parts of the content may be limited by read-only access while other parts may be edited to correct or to supplement data. Thus, we allow to assign read/write properties at the finest granularity level offered by a resource. In case of a data table structured by attributes, modification properties may be set for each attribute separately as well as for each table entry. Consider a clinical data annotation project where a group of medical scientists creates a data table collaboratively. In order to protect table entries that are assumed to be complete from inadvertent modifications, those entries may be set to read-only. In addition, some data that was accumulated from a reliable source (e.g. date of surgery from a clinical information system) is set to read-only to preserve data integrity. The CSCW system has to offer users the functionality to set modification parameters. Sensible standard parameters are essential for the usability. Internally, the system can realise the modification parameters by disabling all users' authorisation to write the data. Consequently, the CSCW system needs to guarantee an exact authorisation method to prevent data from unauthenticated read-access and unauthorised modifications.

4.8 Extension Parameters

Although well-defined data structures ease and standardise the collaborative data entry process, they lack flexibility and extensibility. As collaborative work is an evolutionary process, data may be restructured in order to fit to upcoming demands. We allow two types of data structure modifications: property addition at the conceptual level and property addition at the instance level. A shared data resource may be extended by defining a property template including a set of (property name, property type) pairs. The original view on the resource is augmented with these properties and supplementary information may be entered for each data table entry. Moreover, additional information may be assigned to a single table entry. Consider the follow-up data of one patient when recorded in a different hospital. We want to mark the patient with a 'Follow-up data outside hospital' information. Thus, we require object extension at the instance level.

5 Architecture Overview

The Wasabi framework enabling collaborative work as described in the previous sections is a service-oriented architecture [13]. It bases on the JBoss Application Server (AS) in order to fulfil requirements for enterprise solutions and to provide the scalability and performance needed for a CSCW system which is used in a distributed manner like in the GATiB project. Service orientation is an important characteristic in the context of flexibility, adaptability and maintainability, as already mentioned in requirement $R(4)$. Note that the service orientation of Wasabi can be realised since the underlying JBoss AS is also service oriented. Service orientation is also an essential characteristic for a CSCW system since the data stored might be of arbitrary formats and located in arbitrary repositories. Supporting storing and handling content in a flexible way opens CSCW systems to a wide field of collaboration activities and is no longer limited to specific collaborative environments or applications. The Wasabi's functionality enables on the one hand data to be stored in databases or in file systems reachable via the network. On the other hand it can link already existing databases and data sources to provide users access to those data under consideration of their rights. This is particularly beneficial for the GATiB project since the data collected in different hospitals and institutes is thereby available for the whole community without transferring the data for collaboration to a specific repository. As depicted in Figure 5 the server consists of four main components. The core of the Wasabi enterprise server architecture implements the framework for a CSCW system and aggregates all services of Wasabi (marked with (4) in Figure 5).

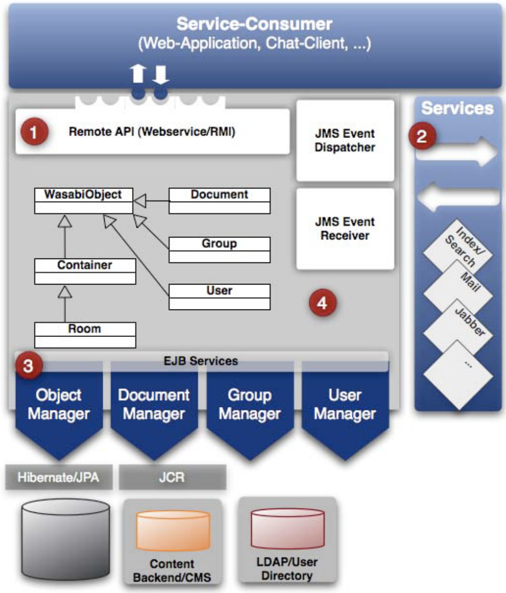

Fig. 5. Main Architecture

The event dispatcher and receiver are responsible for handling internal events, including the arrival of messages, by calling appropriate methods of the internal EJB services (marked with (3)). These services are responsible for handling the objects of the Wasabi core, e.g. the **UserManager** is the service responsible for handling, modifying, and extracting information of **user** objects. These services classify the EJB services as basic services since they implement basic functionalities on the objects of the Wasabi core. Since their tasks focus on the modification and provisioning of data stored in the Wasabi core objects, they can be further classified as data-centric basic services. EJB services are also used to realise flexible user authentication mechanisms as well as to be adaptable to various content backends/repositories.

Third, the Remote API provides an interface for client-server communication (marked with (1)). Therewith a common interface can be used to send requests to different adapted services, in the case the services expect the same input data. This simplifies the enhancement of Wasabi Beans by adapting new services with little effort. The fourth component is responsible for the message exchange with adapted services (marked with (2)). It generates the outgoing messages according to the defined interfaces of the remote webservices and processes and extracts information from incoming messages. After receiving an incoming message, appropriate methods of the Wasabi services are called. Internally individuals, groups, and rooms are objects persistently stored in the central database. Since also the related classes are subclasses of `WasabiObject`, it is necessary to assign to each object an unique identifier (UUID). Storing all `WasabiObjects` in a central database, supports to generate a unique identifier for a `WasabiObject`, i.e. no room or individual can have the same UUID. To provide best possible adaptability, entities of the Wasabi architecture provide a remote API. Main classes are entities like `WasabiObject`, `Document`, `Group`, `User`, `Container`, and `Room`, since these classes realise the data organisation and implement our concept of virtual rooms. To make objects persistent, it is necessary to implement the EJB services (`DocumentManagerService`, `ObjectManagerService`, and `UserManagerService`) as data-centred basic services. The EJB services are data-centred, since they are responsible for data storage and data access. We support different persistency layers and encapsulate this functionality through the EJB services. This encapsulation of functionalities into basic services allows modification on the data access and data storage functionalities without entailing any changes on the server core.

6 Conclusion

In this paper, we presented the main requirements for a CSCW system supporting collaboration in medical environments and outlined a conceptual model and a system architecture fulfilling the specific demands. We found the Wasabi framework capable of supporting the cooperation processes in our project, as it allows the creation and management of knowledge spaces based on a flexible object model and a service-oriented architecture. Though, there is still adjusting work left. We have to integrate data resources into persistency layers, map organisational data, and wrap biomedical applications into services appropriately. We implemented a workbench to access and anonymise distributed data sources. We are working on the extension of the workbench in order to implement the project-centred CSCW client as presented in Figure 2. In our future work we will focus more intensely on service composition in the biomedical context and on formalisation of medical knowledge. That is, the functionality of our GATiB-CSCW will be considerably enhanced and it may be utilised as a virtual platform for collaborative research.

References

1. Armendolia, S.R., Estrella, F., McClathey, R., et al.: Managing pan-european mammography images and data using a service oriented architecture. In: IDEAS-Workshop on Medical Information Systems (2004)
2. Assel, M., Krammer, B., Loehden, A.: Management and access of biomedical data in a grid environment. In: Cracow Grid Workshop (2006)
3. Asslaber, M., Abuja, P., Stark, K., Eder, J.: The genome austria tissue bank (gatib). Pathobiology 74, 251–258 (2007)
4. Bouillon, Y., Wendling, F., Bartolomei, F.: Computer-supported collaborative work (cscw) in biomedical signal visualization and processing. Trans. on Information Technology in Biomedicine 3 (1999)
5. Chu, X., Lonie, A., Harris, P., Thomas, S., Buyya, R.: A service-oriented grid environment for integration of distributed kidney models and resources. In: Concurrency and Computation: Practice and Experience (CCPE) (2007)

6. Arbona, A., et al.: A service-oriented grid infrastructure for biomedical data and compute services. IEEE Trans. on Nanobioscience 6 (2007)
7. GeneSpring. Cutting-edge tools for expression analysis, `www.silicongenetics.com`
8. Li, W., Krishnan, S., Mueller, K., Ichikawa, K., et al.: Building cyberinfrastructure for bioinformatics using service oriented architecture. In: Sixth IEEE Int. Symposium on Cluster Computing and the Grid Workshops (2006)
9. Makris, L., Kamilatos, I., Kopsacheilis, E.V., Strintzis, M.G.: Teleworks: A cscw application for remote medical diagnosis support and teleconsultation. Trans. on Information Technology in Biomedicine 2 (1998)
10. Sweeney, L., Samarati, P.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. In: Proc. of the IEEE Symposium on Research in Security and Privacy (1998)
11. Papazoglou, M.P.: Service -oriented computing: Concepts, characteristics and directions. In: wise, vol. 00, p. 3. IEEE Computer Society Press, Los Alamitos (2003)
12. Sartipi, K., Yarmand, M., Down, D.: Mined-knowledge and decision support services in electronic health. In: Int. Workshop on Systems Development in SOA Environments (2007)
13. Schulte, J., Hampel, T., Bopp, T., Hinn, R.: Wasabi framework an open service infrastructure for collaborative work. In: The third Int. Conference on Semantics, Knowledge and Grid (2007)
14. Stark, K., Eder, J., Zatloukal, K.: Priority-Based k-Anonymity Accomplished by Weighted Generalisation Structures. In: Tjoa, A.M., Trujillo, J. (eds.) DaWaK 2006. LNCS, vol. 4081, pp. 394–404. Springer, Heidelberg (2006)
15. Sturn, A., Quackenbush, J., Trajanoski, Z.: Genesis: Cluster analysis of microarray data. Bioinformatics 18(1), 207–208 (2002)
16. Tusher, V.G., Tibshirani, R., Chu, G.: Significance analysis of microarrays applied to the ionizing radiation response. In: Proc. Natl. Acad. Sci., vol. 98 (2001)