

Open European Nephrology Science Center

Thomas Schrader
thomas.schrader@computer.org

University of Applied Science Brandenburg
Department Informatics & Media
Magdeburger Str. 50
14770 Brandenburg
Charité Universitätsmedizin Berlin
Institute of Pathology
Charitéplatz 1
10117 Berlin

Sabine Hanß

Charité Universitätsmedizin Berlin
Institute of Medical Informatics
Hindenburgdamm 30
12200 Berlin

Sonja Niepage

Charité Universitätsmedizin Berlin
Institute of Pathology
Charitéplatz 1
10117 Berlin

urn:nbn:de:0009-5-33262

Abstract

The Open European Nephrology Science Center is the single center for scientific data in a medical domain supported by the German Research Foundation (DFG). Although the medical domain of nephrology is very small compared to others, e.g. carcinoma, the diseases of this domain have a high impact for the patients and for society. The diagnosis and treatment of these relative rare diseases are a challenge for clinicians, pathologists and other experts. There is a need to improve the data management to enhance the scientific work by support of study recruitment and explorative data analysis.

The Open European Nephrology Science Center represents an organisational structure as a model for scientific collaboration and data collection as well as a web based platform for scientific data management. It covers all aspects of data integration, retrieval and presentation and includes clinical data and Virtual Slides (such called Whole Slide Images (WSI)). Especially two aspects are different to comparable systems: Besides a user management a specialized ticket system covers the resource rights of the data owner. A very flexible database structure based on the RDF model (Resource Description Framework) allows an integration of any kind of resource from any medical domain.

Keywords: e-learning; whole slide images; medical data; data collection; scientific collaboration;

1 Scientific Data Management in Medicine

Day by day a huge amount of patient data are collected by clinicians, pathologists, and other specialists or professionals. They document the diagnostic and treatment process and cover very valuable information about a specific case and from another point of view about the disease itself. Especially for rare diseases it is a need to re-use these data due to the fact that it is quite complicated to collect enough cases and data for studies or explorative data analysis.

In the current literature the topic of secondary use of clinical data is discussed. But every project about data management in the medical domain has to deal with three important illusions:

Standardization All medical data have the same, unique structure.

Data Quality Medical data have a high data quality.

Scientific Community Medical Scientists want to share their data.

In the medical domain two standards are most important: HL7 (Health Level 7) and DICOM (Digital Imaging and Communication in Medicine). Both standards are well established and are used in every hospital information system. But the logical consequence that all data would have the same structure is wrong. Due to different clinical domains, specializations, focus and diagnostic as well as therapeutic strategies, every department has an own structure, content and amount of data. A scientific data management system has to cover at least four main requirements.

- Flexibility & scalability
- Data privacy protection
- Transparency of data usage
- Data quality

The main aim of scientific data management is the allocation of clinical data for a scientific purpose concerning all restrictions of data privacy protection. A scientist with a specific research question or an idea wants to know where are relevant cases for that question. He or she wants to have one point as access. They wouldn't go around to clinical departments and look for handwritten clinical history.

The access of clinical data can be offered in different ways. One strategy is the development of an centralized register for research relevant data (figure 1).

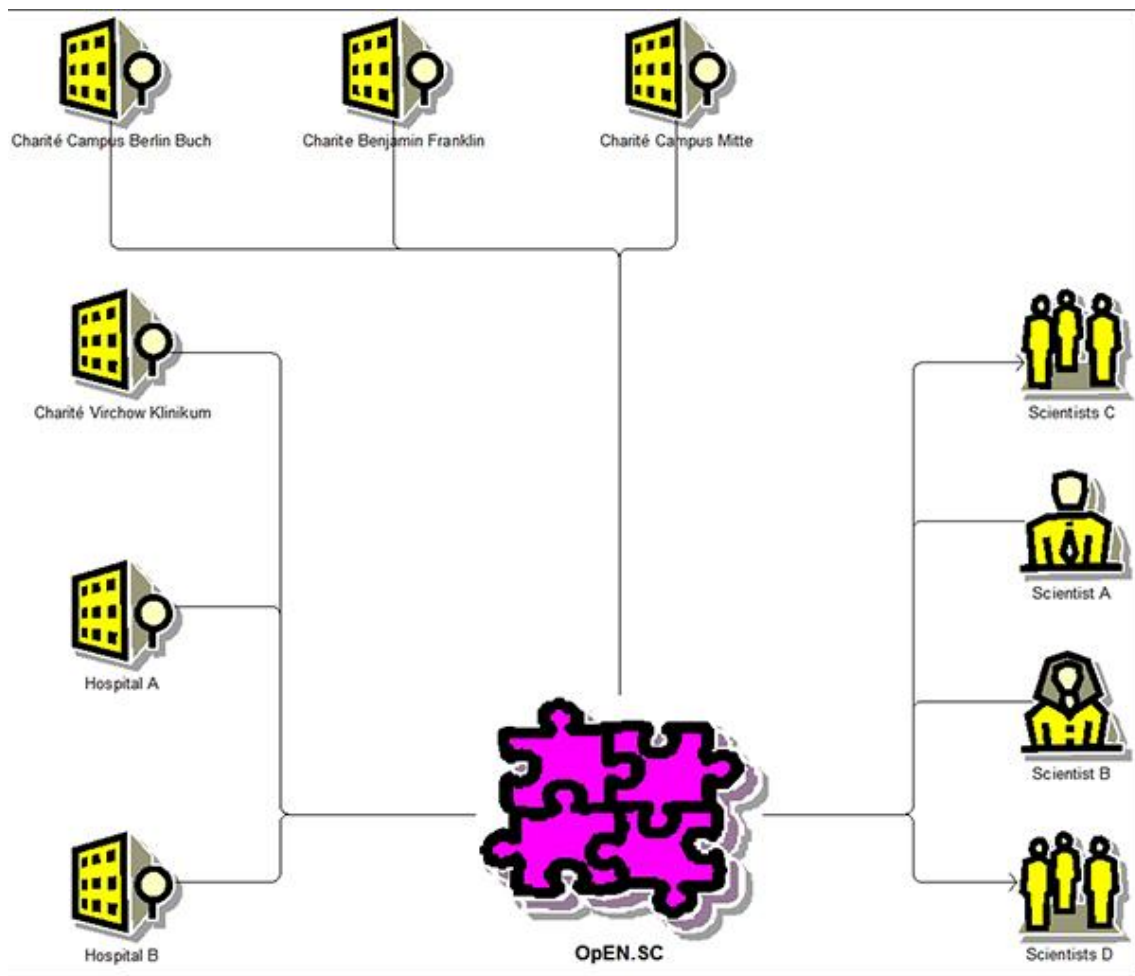


Figure 1: Scientific data management

To manage clinical data under the conditions of data privacy protection in the medical domain the Open European Nephrology Science Center was developed. The German Research Foundation supported this project within the program of Competence Centers for Research Information.

2 Open European Nephrology Science Center - Model, Organization & Principle

The Open European Nephrology Science Center (OpEN.SC) is a webbased platform for data management and an organisational structure (figure 2). As software platform OpEN.SC consists of collaboration and data management tools.

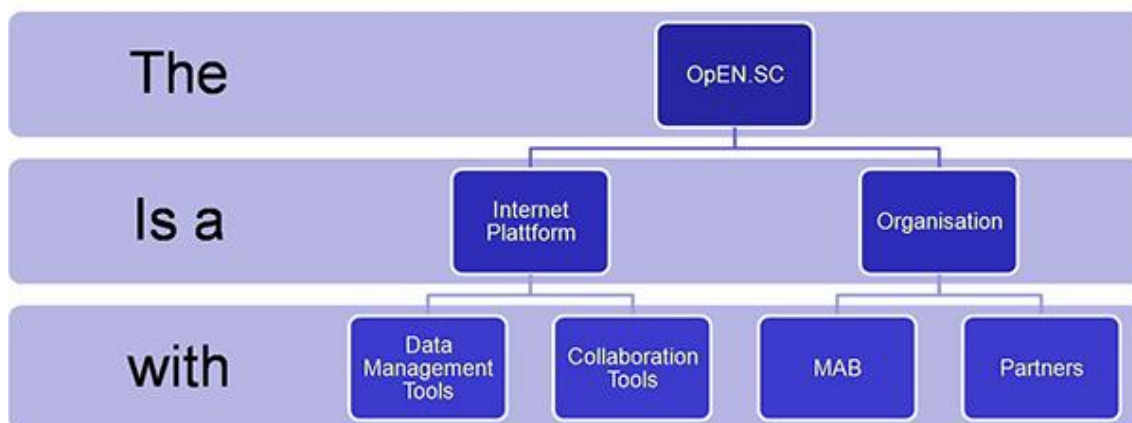


Figure 2: The OpEN.SC Webbased platform and organisational structure

The organisational structure includes a medical advisory board and various project partners. The medical advisory board takes care of the enforcement of the OpEN.SC principles which are:

- Openness
 - New partners are welcome.
 - New projects will be supported by the medical advisory board and their data resources.
- Transparency
 - Every documentation is freely available. OpEN.SC offers the access to the project report and software documentation.
 - The Medical Advisory Board takes care of the fair usage of data in a scientific context.
- Data security and data privacy protection
 - All clinical data are anonymized.
 - A decided user management takes care of the access and availability of data. All processes are traced, all data owners get regular reports about the usage and access of their data.

2.1 Processes of data management

The OpEN.SC platform is based on a Service Oriented Architecture (SOA) to ensure flexibility and scalability. All tasks are modeled as processes and each process consists of web services. Various web services are re-used in different processes. The process view opens the possibility to design the platform for the users coming from the very specific medical domain. The design process can start at a very abstract level discussing with experts and users.

For the scientific data management various processes were identified (figure 3). All processes cover all aspects of data management and consist of the processes of data input, retrieval and presentation and include the management of user related to the data and resources. Beside the main data management processes, image handling is included: the scanning of virtual slides (such called Whole Slide Images), image retrieval and image analysis.

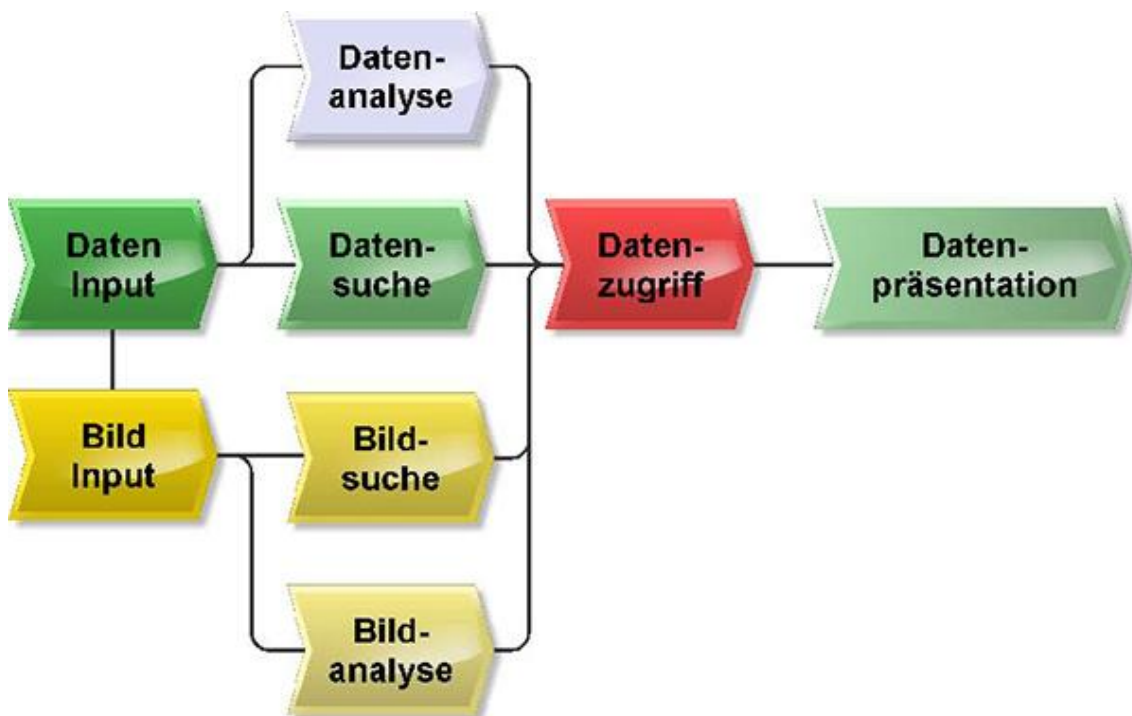


Figure 3: Processes of data management

The processes are realized by web services. The system has web services which can be divided into (figure 4):

- Infrastructure services, which cover the basic functionality: database, portal etc.
- Identity Management webservices for authentication and authorization of actions within the platform
- Data retrieval services for retrieval and presentation of clinical data
- Data input services for data import.

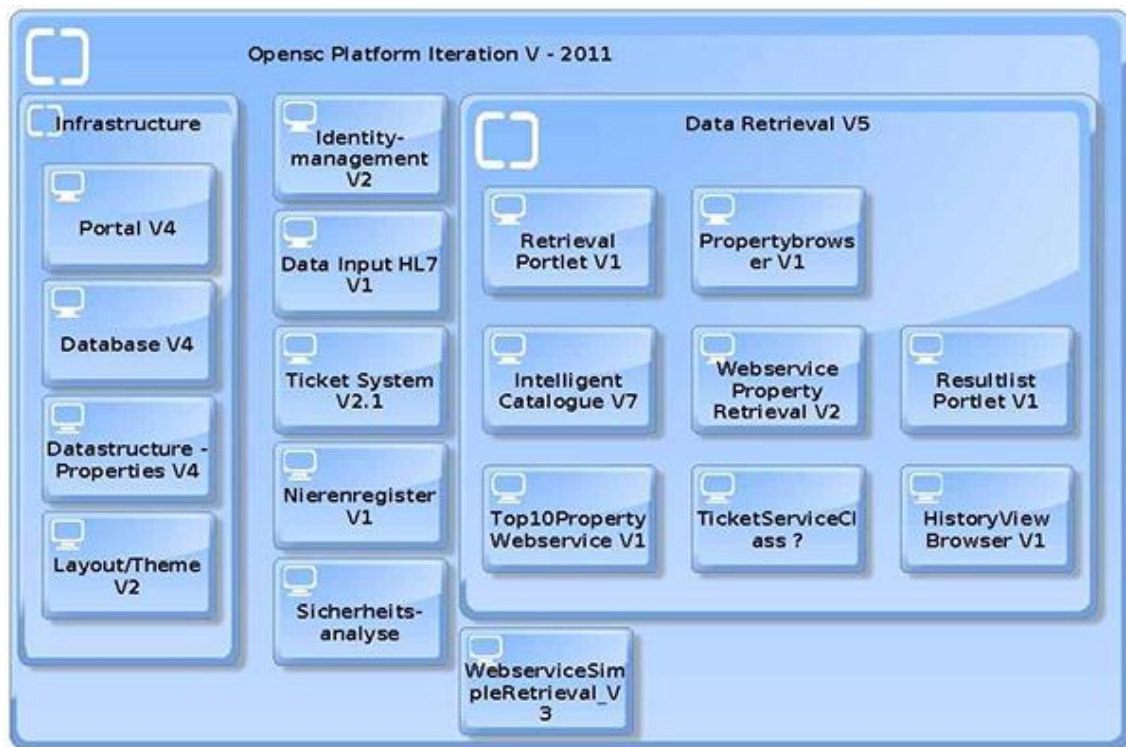


Figure 4: Overview about the web service structure of OpEN.SC based on SOA

2.2 Partner & data integration

For distributed databases different models of data management exist:

Repository Model All data are stored in a centralized repository. A query will be sent from a user directly to the repository. The data resources do not know anything about the further usage and cannot control the access to their data directly.

Federated Database The data are stored in peripheral separated databases. A specialized service manages queries entered by a central user interface. This service adapts the query for the different data resources and collects the data from them.

In the medical domain a federated database model is not possible to implement due to legal issues (data privacy protection) and risks of Internet attacks. Usually it is not allowed to send a query directly from outside to a hospital information system. The single option is a repository where data will be sent actively and automatically from resources to the repository after a special data preparation process for transmission of encrypted and anonymized data.

This has consequences for data management as well as user management. The user management is not only limited to the end user but also to the resources. The resources want to keep in contact with their own data. The access of the data can be controlled by the data owner.

The important role of the data owner over the complete lifetime of the data is covered by the process of integration of new partners into the repository (figure 5). A new clinical department can send a representative to the Medical Advisory Board as central organisational structure. A contract defines any rules and responsibilities for data transfer and management. It includes a definition what kind of data and the amount of data can be sent to the repository. The electronic version of this document is used as property file to control the web service for data selection, encryption and transmission. A very difficult problem is the development of a flexible, scalable and adaptable data structure for such a repository. OpEN.SC implemented a triple structure based on Resource Description Framework (RDF).

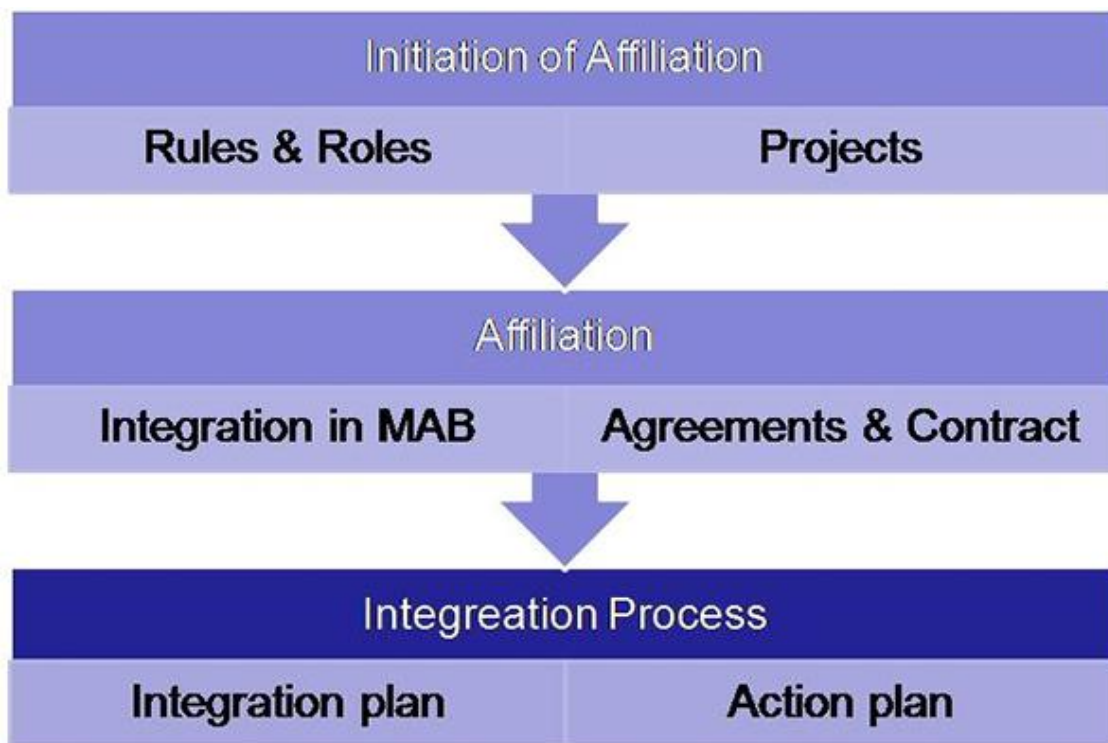


Figure 5: Integration of a new partner based on a contract

The data from various resources will have a very variable data structure. It is impossible to re-design the database structure again and again. The transformation of any hierarchical structure to a triple model (resource - description - value) allows us to load the data very quickly into the system and model the content in an data ontology. Now all data are divided into different domains and are stored in a triple manner or in a Entity-Relation-Model if the data will have a very stable structure e.g. project data. The database model is now called hybrid Domain Model of OpEN.SC (figure 6).

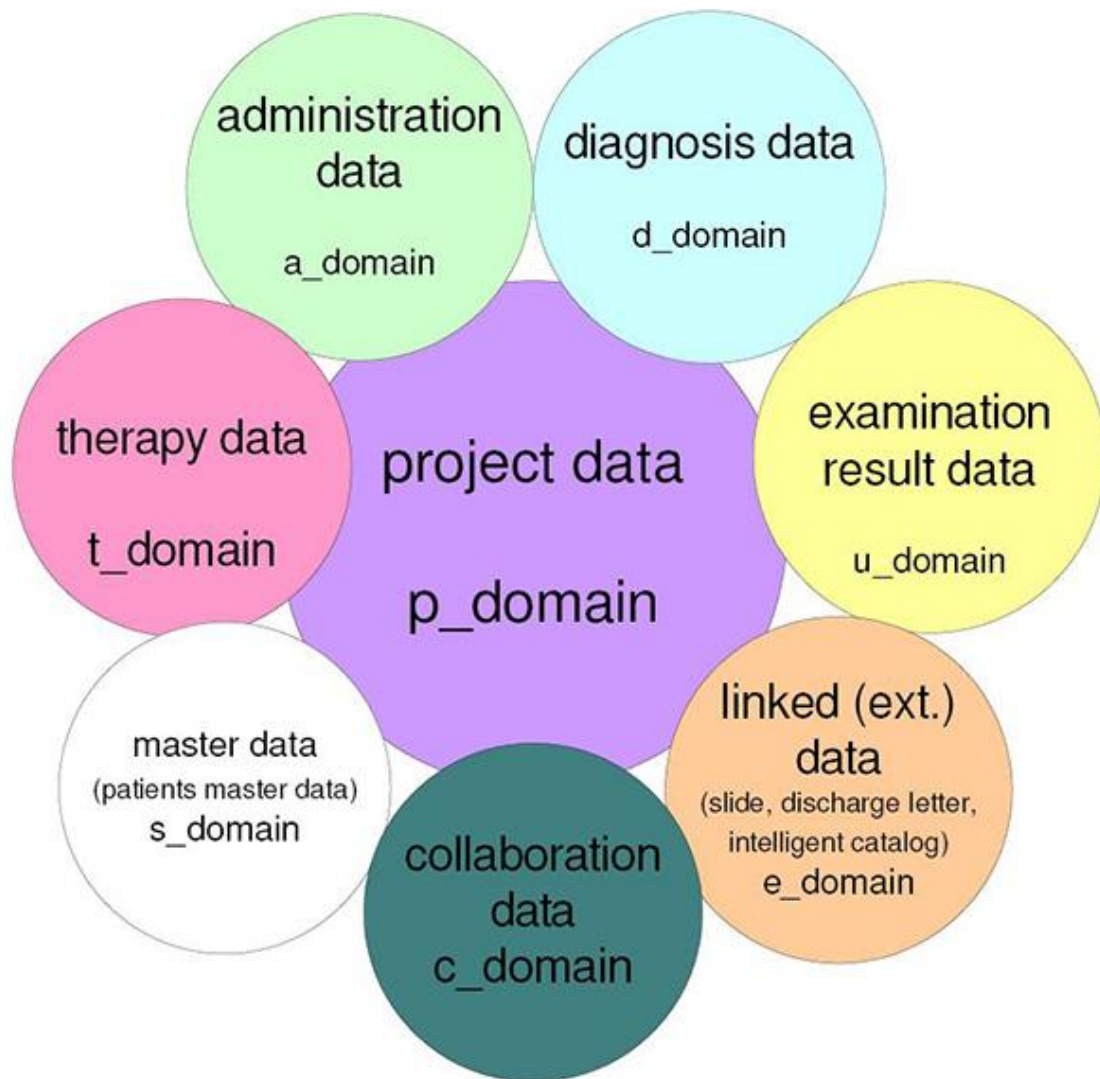


Figure 6: Hybrid Domain Model of OpEN.SC

2.3 Data Retrieval

For data retrieval two different strategies were developed:

- Based on a Case Retrieval Net (figure 7) the user can look for specific cases and as well as for similar cases expressed by distances on case retrieval net graphs.
- The database ontology expresses relations between terms and content and can be used for a dynamic selection of properties in a tree of terms and relations.

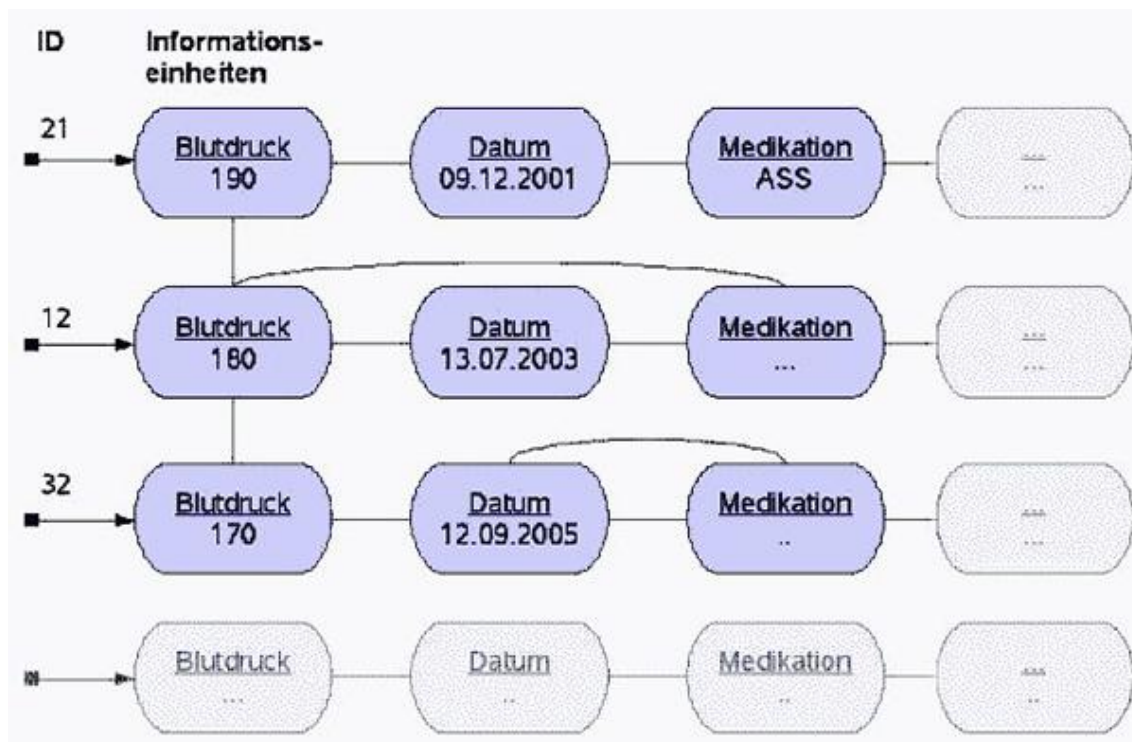


Figure 7: Case Retrieval Net

Although scientists have very specific requests to the database and the property based retrieval, data requests usually can be fulfilled and the database is flexible against changes, such as the inclusion of new partners with new attributes.

2.4 OpEN.SC-Tools

Beside the original clinical data, image information plays a very important role in scientific work. Any kind of images can be linked to a case - X-rays, ultrasound etc. For the domain of nephropathology the images of biopsies are the basis for diagnosis and treatment and therefore quite interesting for scientific evaluations.

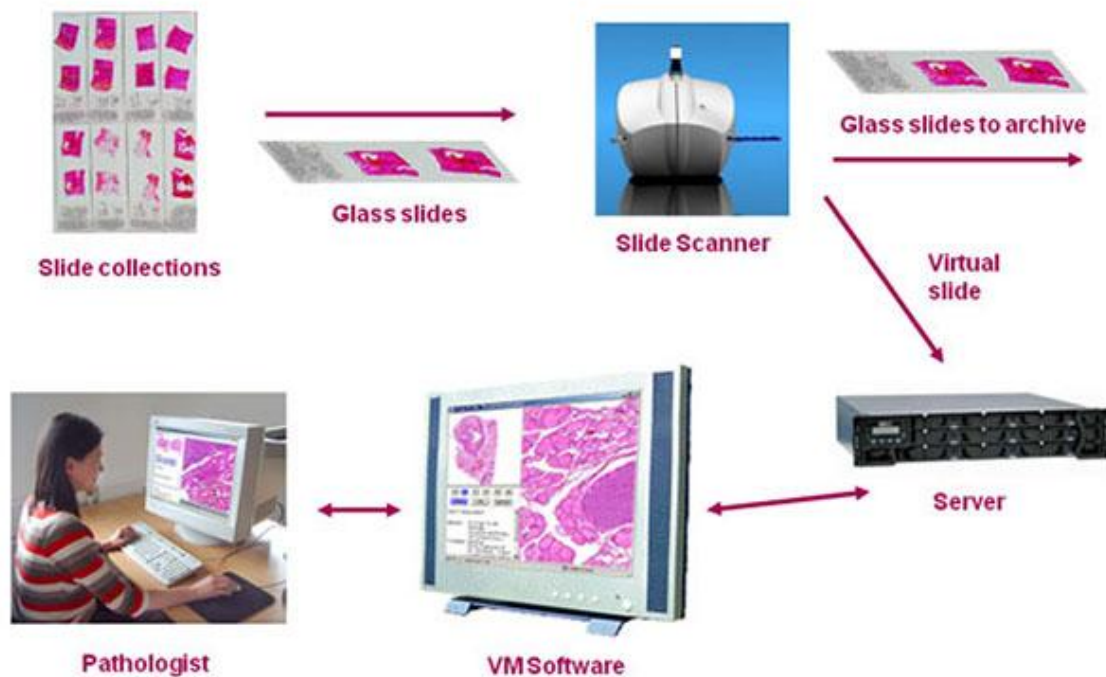


Figure 8: Process of Virtual Microscopie using Whole Slide Images

The glass slides of biopsies in the pathology department can be completely digitalized. The result is so-called Whole Slide Image (WSI). Each WSI can have a file size of 2 GB and special viewing software is used to review the case at a monitor. The management of digital slides is covered by an own process (figure 8). Now the situation is very comfortable. A case stored in OpEN.SC has digitalized clinical data and related WSI of pathology department. The scientist can retrieve or review data as well as images (figure 9). This is the basis for application of data analysis tools as well as image analysis tools. Various commercial tools will be integrated in OpEN.SC for image analysis.

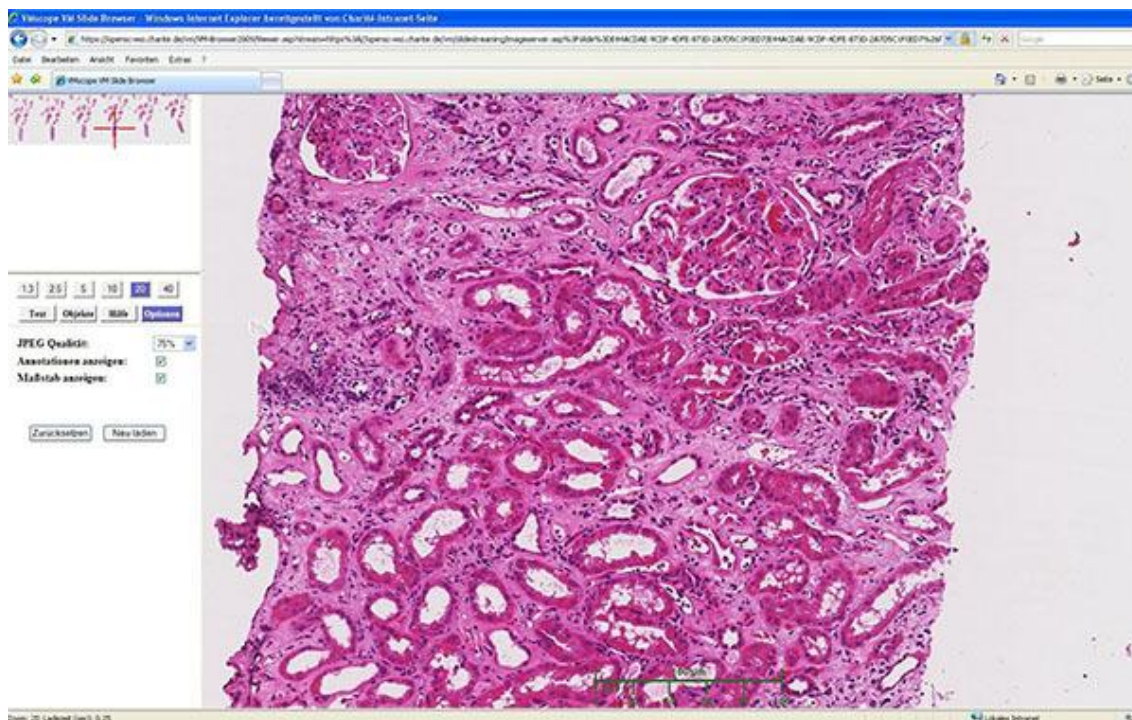


Figure 9: Viewing Whole Slide Image (WSI), such called Virtual Slide

3 Results & Activities

OpEN.SC collects regular data from three clinical departments (table 1). Cases with nephrological disease have a long clinical history. Inflammation processes at the kidney have a damage of the parenchym as a result. At the end of the negative development the kidney will loss his function. Patient has to go to dialysis and in best cases will get a new organ. After transplantation the organ should be treated against rejection for many years.

Dep.	Pa-tients	Items Ad-min	Items Dia-gn	Items Therap.	Items Eval
CVK	5.581	100.863	288.961	1.351.691	41.500.563
CCM	3.386	92.869	232.566	1.277.379	34.286.984
CBF	1.118	27.096	27.647	38.369	4.352.947
Total	10.085	220.828	549.174	2.667.439	80.140.494

Table 1. OpEN.SC Statistics: Cases, data about the cases related to administrative, diagnostic, therapeutic and evaluation domain



Figure 10: OpEN.SC workshop with user tests and system presentation

The long history, the numerous related diseases to renal insufficiency are the reason for a complex patient history with many and quite different types of data. Images are a very important component of clinical information.

All members of the clinical departments have an access to the OpEN.SC repository. In various research projects data are exchanged especially WSI due to the importance for diagnosis and treatment. Regular OpEN.SC workshops took place with experts, users and scientist from various countries (Poland, France, Spain, Lithuania, China, Iran, see figure 9). The workshop was a quite good possibility to embed the project in an international context and get relationships to different research groups.

At these meetings the OpEN.SC system was presented and the user could test the environment, comment the interface and functionality and give advices for future projects and developments. The Open European Nephrology Science Center is an active and living center for scientific data management. The concept of data management allows to add completely different medical domains to this service and allows to retrieve data from different resources and domains. The importances of data privacy and the need for many data sets are not contrary related if the data management system covers the interests of the patient, the scientist as well as the data owner.



Figure 11: Discussion with international Expert for Nephrology

4 Acknowledgment

We would like to thank the German Research Foundation (DFG) for their support and attendance, esp. Dr. Eckelmann.

5 Literature

A list of publications is available from the authors.