

# How can clinical biobanks and patient information be adapted for research – Establishing a hospital based data warehouse solution

Wenche Reed, Sissel Jor and Roger Bjugn

*Oslo University Hospital, Oslo, Norway*

Correspondence: Wenche Reed, Oslo University Hospital, Kirkeveien 166, NO-0407 Oslo, Norway  
E-mail: wre@ous-hf.no Telephone: +47 950 81 804

## ABSTRACT

The use of human biological material and related clinical information is central in basic research, translational research and clinical research. Extracting this information from hospital information systems is time-consuming and labor intensive. Establishing a data warehouse is a possible solution to extract the information in an efficient and secure way. An information model for a data warehouse has been developed at the Oslo University Hospital in collaboration with Akershus University Hospital. Being hospitals, the overall information model is patient-centered. The model also allows for appropriate data exchange between different organizations. The model is described with a focus on some of the major conditions for success with respect to the legal and organizational framework of Norwegian hospitals. A web-based information and tracking system for clinical biobanks is an important element of the model, and one of the aims of the national project Biobank Norway.

## INTRODUCTION

The use of human biological material is paramount in basic research, translational research, and clinical research. The drive towards personalized medicine also depends upon the availability of high-quality biological samples from individual research participants (1). However, clinical and translational research with development of new therapeutic strategies cannot take place with only biological samples; one must also have access to the related clinical information.

In this context, almost all clinical information is primarily generated at hospitals. Norwegian hospitals have been collecting biological samples and related clinical information for ongoing diagnostic and treatment purposes since the 1920's, resulting in millions of identifiable samples currently being stored. One example is the more than 900 000 cell and tissue samples taken annually and stored at pathology laboratories, another is the 210 000 samples taken annually from blood donations and stored at microbiology laboratories for at least 2 years. Despite this rich resource of biological samples and related clinical information, the research turn-out based on this material is not reflective of the great amount of resources invested to collect these biospecimens and related data.

In 2007, two Norwegian ministries commissioned a special report on biobanks and health registries from the Research Council of Norway. The report was published in 2008 and recommended the establishment of a long-term, national platform for medical research (2). One of the specific recommendations was to establish infrastructure facilitating the use of human biological material and clinical data from hospitals. In 2010 a national consortium, Biobank Norway, was granted EUR

10 million from the Research Council of Norway to establish such a national biobank platform. This discrepancy between resources invested in collecting human biological material and the scientific turn-out is not solely a Norwegian phenomenon. A European strategy document on research infrastructure published in 2006 contained a recommendation to establish a pan-European network of biobanks and biomolecular resources (3).

## CURRENT CHALLENGE – CLINICAL DATA AND SAMPLES ARE NOT READILY AVAILABLE FOR RESEARCH

All individuals in Norway are assigned a unique 11-digit identifying code that is used in the public health and social care sector and in all public registries. Norwegian hospitals use electronic health records and therefore it is possible to identify and track every individual through the health care system. However, all health care information systems have traditionally been developed with the aim of identifying a particular patient, not to identify groups of patients and extract their data or samples for research purposes. Most hospitals have several hundred different specialist information systems, and the systems have not been designed to communicate with one another using a common standard. In addition, primary data are quite often presented in an unstructured format. Consequently, even if the information is available in electronic format, it is very time-consuming and labor intensive to extract data for quality assessment and research purposes. This challenge is of course not unique for Norway. DeWitt and coworkers have claimed that the academic health system is “rich in data and poor in information” (4).

The accessibility and renewed use of samples in the different research biobanks in hospitals are also hampered by lack of hospital-wide tracking systems, leaving individual researchers to create their own local systems that are not accessible to others.

### POSSIBLE SOLUTION – THE DATA WAREHOUSE

In order to extract existing electronic data for secondary quality and research purposes in a timely, efficient and secure way, several organizations propose the establishment of a data warehouse (4-8). A data warehouse can be defined as “a copy of transaction data specifically structured for query and analysis” (see ref. (7)). In this context, transaction data refers to patient data continuously collected as part of ongoing clinical care provided by the hospital. A hospital based data warehouse must be designed so that it does not interfere with daily operations (e.g. not compromise concurrent clinical care processes), and it must be flexible in its overall design so that upgrades and replacements of primary information systems will not be affected.

### AN EXAMPLE FROM NORWAY

Since 2009 Oslo University Hospital has been collaborating with Akershus University Hospital on a data warehouse concept. To ensure adequate organizational support, the development was approved by the CEOs of both hospitals. The concept has evolved in close collaboration with the hospitals’ data protection office, IT-architects, relevant organizational units and researchers. The concept development was based on the following leading principles:

- Protection of personal data must conform to national and European regulations
- Data security measures must conform to national regulations and recommendations
- Data storage must be based on a predefined legal basis (patient treatment or research)
- Data architecture must be based on national recommendations and be compatible with recognized international standards for data exchange and nomenclature
- The data warehouse solution should be acquired through a public tender and in-house development should be minimized

Being a hospital, the overall information model chosen for the data warehouse was patient-centered (Fig. 1).

The clinical data warehouse design itself (Fig. 2) reflects the different legal provisions for acquisition and the use of personal health information in Norway. The aim of the infrastructure is to establish a platform or crosswalk to connect clinical data and research data within the existing legal framework.

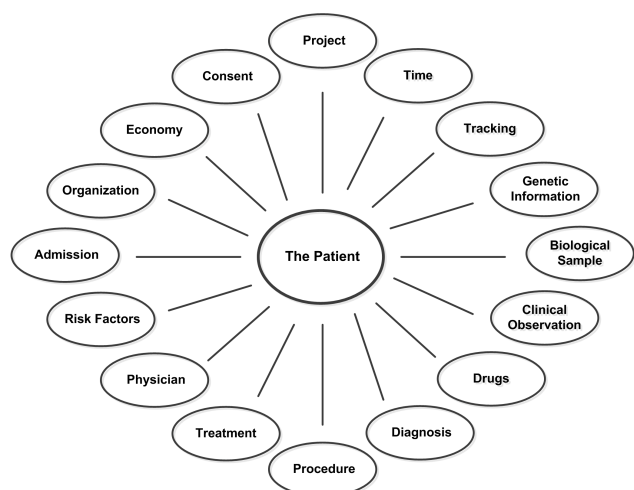
According to Norwegian legislation, acquisition and storage of personal health information in the health care system is based upon assumed consent. Such personal health information can be stored indefinitely. The data

can be used for quality assurance and quality improvement studies without consent. In contrast, acquisition and storage of research data requires explicit consent or a waiver of consent granted by an ethics committee. Such research data cannot be stored indefinitely.

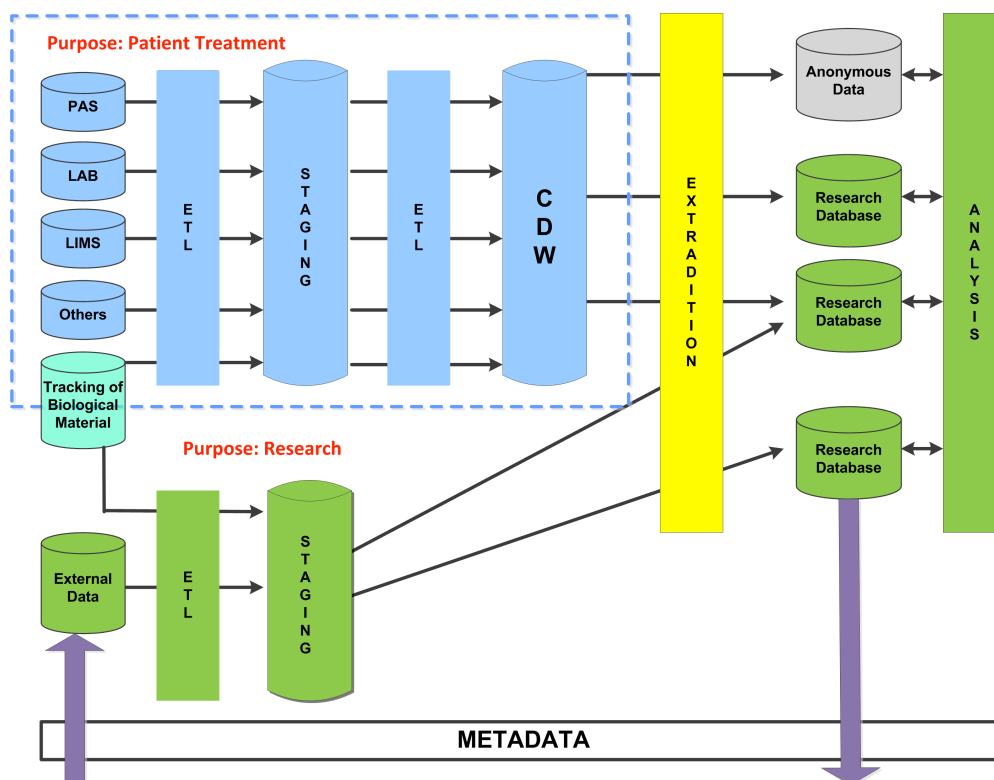
The data warehouse model was therefore designed to distinguish clearly between data obtained from patient care and data generated by research projects. The main part of the clinical data warehouse consists of clinical data (inside the box with dashed lines in Fig. 2). Data are extracted from the different clinical systems, cleansed and transformed in the staging area, and finally loaded into the clinical data warehouse. De-identification is performed in the staging area before the loading process to ensure patient privacy. The data in the clinical data warehouse are in a structured form and available for data mining or analyzes through an honest broker mechanism (marked Extradition in Fig. 2) (see (9,10)). An honest broker in this context is a person who has the authority and means to ensure proper handling of personal data. This function will be undertaken by the data protection office and will to a large degree be based upon automatic procedures. When a researcher has the appropriate ethical approval and patient consent, the honest broker can extract the relevant data from the central clinical data warehouse to the appropriate research database (marked research database in Fig. 2).

An important part of the concept is to establish an institutional information system for the tracking of biological samples in the hospital (see Fig. 2). A tracking system will have a dual purpose; tracking both diagnostic samples and samples obtained for research. This is the only way to enhance the scientific use of the hospitals biobanks. Establishing a tracking solution at the institutional level is one of the aims of the national Biobank Norway project.

The model will also facilitate data exchange between the different organizations, i.e. between population cohort biobanks and hospitals, thereby fulfilling one of



**Figure 1.** Patient-centered information model for the data warehouse.



**Figure 2.** Model for data warehouse.

PAS: Patient administrative system; LAB: Laboratory information systems; LIMS: Laboratory information management system; ETL: Extraction, transform, and load; CDW: Clinical Data Warehouse.

the goals of Biobank Norway. As the clinical data warehouse is based on information from patient care, import of external research data from other hospitals or population cohorts will have to be stored in the research data (marked External data in Fig. 2). However, similar rules for cleansing, transformation and de-identification as in the clinical data warehouse will be applied on research data.

**DISCUSSION**

The data warehouse model described will support both clinical care and clinical and translational research. More specifically, it will facilitate the exchange of biological material and personal health information between the research organizations participating in the Biobank Norway consortium. Despite encompassing all primary information systems present in large university hospitals, the model is so general and flexible that it can be employed in any hospital, independent of size, and accordingly be used both if more hospitals were to join the Biobank Norway consortium, and if hospitals independent of Biobank Norway want a common platform for information exchange.

The data warehouse model as such shares many similarities with other models (4-8). As noted by Lyman and colleagues, the design of an integrated data warehouse is a resource-intensive process requiring a multi-disciplinary approach and substantial investments of time and energy (7). However, in our experience there

are two other issues which are just as important as technical aspects and resources required.

**Context and external stakeholders**

Developing a data warehouse in a private organization is very different from developing a similar solution in a public Norwegian hospital. All Norwegian hospitals are organized into four regional health authorities, which again are controlled by the Ministry of Health and Care Services. Accordingly, no major development in a hospital can take place unless (officially or unofficially) approved by the hospital owner(s). In addition, the four regional health authorities and some public bodies have established a common body (National ICT; <http://www.nasjonalikt.no/>) for the coordination of information and communication technologies in hospitals. This means that all hospitals have to abide to the general ICT-policies laid down by the National ICT. Realizing this project environment, a stakeholder analysis was undertaken in the early stage of our project in order to identify major stakeholders. During further project development these stakeholders were formally addressed in order to reduce project risk as much as possible.

**Security and confidentiality**

The legal framework for the acquisition and storage of personal health information is one of the key elements in the design of a hospital based data warehouse. As in most countries, Norwegian legislation is very specific

and strict with respect to the handling of personal health information. In our project, we chose to look upon these regulations as opportunities, not as burdensome hindrances against research. The hospital has a formal responsibility to establish infrastructure and routines for ensuring continuous quality assessment and improvement of ongoing patient treatment. The establishment of a clinical data warehouse solution (inside the box with dashed lines in Fig. 2) will just facilitate this and can accordingly be undertaken within an already existing framework. However, such a warehouse might also increase the risk of misuse of personal health information as more sensitive data are gathered (and searchable) in one place. It was therefore early on decided that the organization had to establish a dedicated and impartial gatekeeper to the data warehouse. The data protection office, with its already established formal mandate and duties, were identified as the organizational unit best suited for this task. As the proposed model to a large extent is based upon the use of clinical data, the data protection office recommended that the formal “ownership” of the clinical data warehouse should be the person in charge of medical services/quality assessment. This decision is still awaiting formal approval by the hospital’s CEO.

Our project has experienced some additional challenges. Oslo University Hospital is a merger between three previous hospital trusts, and this merger has taken place at the same time as this project was to be developed. This has caused a number of time-delays as the new management structure had to be put in place before major decisions could be taken. In addition, the regional health authority decided to outsource hospital

ICT-services to a separate ICT-service provider in the same time period. This process also caused a number of practical and formal challenges.

Despite this, the project has managed to move forward. A tender for the acquisition of a software system for the tracking of biological material has recently been made public. The next step is to undertake a proof-of-concept study with some of the primary patient information systems.

## CONCLUSION

The use of human biological material and related clinical information is central in both basic research, translational research and clinical research. Extracting this information from the information systems in the hospitals is very time-consuming and labor intensive. Establishment of a data warehouse is a possible solution for extracting the information in an efficient and secure way, and access the clinical biobanks. Therefore an information model for a data warehouse has been developed at the Oslo University Hospital in collaboration with Akershus University Hospital. The model also allows for appropriate data exchange between different organizations, and will as such facilitate scientific collaboration between population based health surveys and the hospitals.

## ACKNOWLEDGMENTS

This research was supported through funds from Biobank Norway – a national infrastructure for biobanks and biobank related activity in Norway – funded by the Norwegian Research Council (NFR 197443/F50).

## REFERENCES

1. Hewitt RE. Biobanking: the foundation of personalized medicine. *Curr Opin Oncol* 2011; **23**: 112-119.
2. Gode biobanker – bedre helse (Good biobanks – better health). Oslo: The Research Council of Norway), 2008, p. 1-93.
3. European roadmap for research infrastructures: Report 2006. Luxembourg: European Strategy Forum on Research Infrastructures (ESFRI), 2006, p. 1-86. ([http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri-roadmap&section=roadmap-2006](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap&section=roadmap-2006))
4. DeWitt JG, Hampton PM. Development of a data warehouse at an academic health system: knowing a place for the first time. *Acad Med* 2005; **80**: 1019-1025.
5. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc* 2010; **17**: 131-135.
6. Hu H, Correll M, Kvecher L, Osmond M, Clark J, Bekhash A, et al. DW4TR: A data warehouse for translational research. *J Biomed Inform* 2011; **44**: 1004-1019.
7. Lyman JA, Scully K, Harrison JH, Jr. The development of health care data warehouses to support data mining. *Clin Lab Med* 2008; **28**: 55-71, vi.
8. Rajala T, Savio S, Penttinen J, Dastidar P, Kahonen M, Eskola H, et al. Development of a Research Dedicated Archival System (TARAS) in a University Hospital. *J Digit Imaging* 2011; **24**: 864-873.
9. Boyd AD, Saxman PR, Hunscher DA, Smith KA, Morris TD, Kaston M, et al. The University of Michigan Honest Broker: a Web-based service for clinical and translational research and practice. *J Am Med Inform Assoc* 2009; **16**: 784-791.
10. Dhir R, Patel AA, Winters S, Bisceglia M, Swanson D, Aamodt R, Becich MJ. A multidisciplinary approach to honest broker services for tissue banks and clinical data: a pragmatic and practical model. *Cancer* 2008; **113**: 1705-1715.