SPIRIT: Systematic Planning of Intelligent Reuse of Integrated Clinical Routine Data

A Conceptual Best-practice Framework and Procedure Model

W. O. Hackl; E. Ammenwerth

Institute of Biomedical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

Summary

Background: Secondary use of clinical routine data is receiving an increasing amount of attention in biomedicine and healthcare. However, building and analysing integrated clinical routine data repositories are non-trivial, challenging tasks. As in most evolving fields, recognized standards, well-proven methodological frameworks, or accurately described best-practice approaches for the systematic planning of solutions for secondary use of routine medical record data are missing.

Objective: We propose a conceptual best-practice framework and procedure model for the systematic planning of intelligent reuse of integrated clinical routine data (SPIRIT).

Methods: SPIRIT was developed based on a broad literature overview and further refined in two case studies with different kinds of clinical routine data, including process-oriented nursing data from a large hospital group and high-volume multimodal clinical data from a neurologic intensive care unit.

Results: SPIRIT aims at tailoring secondary use solutions to specific needs of single departments without losing sight of the institution as a whole. It provides a general conceptual best-practice framework consisting of three parts: First, a secondary use strategy for the whole organization is determined. Second, comprehensive analyses are conducted from two different viewpoints to define the requirements regarding a clinical routine data reuse solution at the system level from the data perspective (BOTTOM UP) and at the strategic level from the future users perspective (TOP DOWN). An obligatory clinical context analysis (IN BETWEEN) facilitates refinement, combination, and integration of the different requirements. The third part of SPIRIT is dedicated to implementation, which comprises design and realization of clinical data integration and management as well as data analysis solutions.

Conclusions: The SPIRIT framework is intended to be used to systematically plan the intelligent reuse of clinical routine data for multiple purposes, which often was not intended when the primary clinical documentation systems were implemented. SPIRIT helps to overcome this gap. It can be applied in healthcare institutions of any size or specialization and allows a stepwise setup and evolution of holistic clinical routine data reuse solutions.

1. Introduction

We are currently facing an exponential growth of information in the healthcare sector. This will pose a range of challenges in various domains, especially in the health informatics field [1]. On the other hand, this "information explosion" can open up undreamed-of possibilities of secondary use, defined as the (re)use of clinical routine data outside direct care for multiple purposes [2]. Clinical routine data are considered extremely precious [3, 4] and secondary use can be beneficial for almost all stakeholders in the healthcare sector, including policymakers, public health officials, scientists, clinicians, citizens, and industry. But of course, secondary use implicates new challenges and problems [5]. In recent years, therefore, the International Medical Informatics Association (IMIA) has devoted consistent activity to tackling these reuse problems on an international level [5]. And in the field, a multitude of excellent examples of clinical data warehouses, routine data reuse projects, frame-

works, and initiatives can be found, e.g. [6–17].

Recently, the term *big data*, which refers to a set of new technologies for managing, processing, and analysing massive and disparate data sets [18], has been receiving increased attention in biomedicine and healthcare [19]. However, clinical routine data are not necessarily "big" in terms of volume but are instead quite challenging in terms of variety. Thus, building and analysing integrated clinical routine data repositories are nontrivial, challenging tasks [20, 21]. There are several reasons for this. First, the clinical context is very complex and clinical processes involve multiple, multi-disciplinary, hierarchically organized healthcare professionals from different organizations. Second, most of the clinical tasks are not standardized. A good portion of them is carried out collaboratively and a few decisions can be distributed over various stakeholders. Third, methods of clinical documentation vary depending on the documentation purpose and resulting data are often semi-structured or unstructured. Even if documentation is structured, in many cases, it is not standardized or follows proprietary, institution-specific standards. Fourth, information system architectures are as complex as they are heterogeneous. Usually, they constantly evolve over time. The information systems themselves are of a sociotechnical nature and comprise numerous subsystems with up to hundreds of application systems. Secondary use of clinical data was often not intended when many of these application systems were designed, with the effect that data, once documented and stored in the depths of proprietary database schemes, cannot be retrieved easily. Thus, solutions for reusing clinical routine data are not available out of the box. They must be tailored to the specific context. This might be an explanation for the fact that almost all of the prominent data reuse examples are situated in "lighthouse locations", large healthcare institutions or organizations having sophisticated IT landscapes and highly specialized IT departments with highly skilled staff.

Almost 20 years ago, Leiner and Haux postulated – with good cause – that clinical documentation has to be planned system-atically, and that the purpose for the documentation and collection of each data element has to be defined in advance, with a view to the documentation goals [22]. The idea of secondary use somehow contradicts this postulate because the relevance of single routine data elements cannot be known a priori, as many questions do not emerge before a critical mass of data is available and first investigations point in a certain direction. Conceptual frameworks, procedure models, or well-described best-practice approaches for planning, establishing, and effectively analysing such routine data repositories are hard to find. This makes it very difficult to systematically plan solutions for reusing clinical routine data. In contrast, a "data leech" approach – collecting any available data crumb without knowing how the collected data can or should be analysed – is just as disadvantageous.

Clinical documentation systems are purposive. They are designed to support specific information needs of specific users in specific situations and it is obvious that a system designed to support specific decisions is different from a system that should collect relevant data for future use [23]. This situation leads to a gap between a goal-oriented approach and an approach that leaves scope for development and evolution when designing clinical routine data reuse systems.

2. Objective

To overcome this gap between goal-orientation and scope for evolution, we propose a conceptual best-practice framework and procedure model for the *systematic planning of intelligent reuse of integrated clinical routine data (SPIRIT)*. SPIRIT can be applied in healthcare institutions of any size or specialization. It was developed and refined in two case studies with different kinds of clinical routine data, with process-oriented nursing data, and with high-volume multimodal clinical data from a neurologic intensive care unit.

3. Methods

3.1 Literature Overview: Building the Backbone of SPIRIT

We started with a broad literature overview aiming to scan the state of the art in building and analysing integrated data repositories with a special focus on the intended clinical scope of application. This literature overview was not intended to be systematic in terms of reproducibility. In fact, it was intended to identify a possible best-practice framework for enabling secondary use of clinical routine data and to identify possible application areas for clinical routine data reuse. The overview included scientific literature retrieved from PubMed, Ebsco Business Source Premier, Google Scholar and conference proceedings of the last few years (e.g. MedInfo, MIE, AMIA) as well as textbooks on data warehousing and business intelligence.

We started with general search terms such as "secondary use", "reus*", "re-us*", "data warehous*", "data mart*", "business intelligence", "clinical data", "routine data", "nurs* data", etc. The retrieved publications were then classified into three groups: i) "irrelevant, reject", ii) "partly relevant, analyse abstract only", or iii) "highly relevant, analyse full text". The relevant articles and passages from textbooks were scanned for possible areas of application for clinical routine data reuse, which were collected in a list. The entries in this list were then clustered inductively. The resulting compilation of possible reuse goals was used to generate ideas in the following case studies. The relevant articles and text passages were also scanned for additional search terms, which were then used for additional search runs in the defined databases and on the web. References of articles in group iii were further combed through in a snowball search to identify further relevant literature.

Based on the relevant literature that was found, we attempted to distil a first, raw conceptual best-practice framework for the systematic planning of a routine data reuse solution. For this purpose, we first dissected the different developmental phases of a routine data reuse solution: i) project planning and management, ii) systems analysis and assessment, iii) functional

specification and design, iv) hardware and software implementation, v) system initial deployment, vi) project support, vii) close-out, and viii) operation and lifecycle management. We then identified required tasks to be carried out in these different phases and determined the execution planning for these tasks.

This execution plan, the backbone of SPIRIT, was then used as a procedure model in a first case study, in which a reuse solution for routine nursing data in a large Austrian hospital group was planned and implemented. In this case study, in order to improve and refine the raw SPIRIT framework and to elaborate a preferably detailed best-practice procedure model, we also wanted to learn how such planning and implementation could be done in an optimal way.

3.2 Case Study I (Nursing Data): Development of SPIRIT's Skeletal Structure

The first study was designed as a comprehensive, single case study [24] in the Tyrolean Federal Hospitals (tirol kliniken – formerly TILAK), a large Austrian hospital group. Tirol kliniken comprises the Innsbruck university hospitals plus several district and specialty hospitals with a total capacity of 2,465 beds and 7,102 full-time equivalent (FTE) healthcare professionals and administrative staff members (among them 2,470 FTE registered nurses) who handle approximately 120,000 inpatient and 1,150,000 outpatient stays per year (key figures from 2013).

In 2006, an electronic nursing documentation system (based on the *Cerner Millennium©* clinical information system) was implemented at the Innsbruck university hospitals. Since its initial implementation, the nursing documentation system has been rolled out in almost all of the group's hospitals, resulting in a large amount of mostly structured nursing data from the inpatient and outpatient areas now available in electronic form. The nursing management board realized that these data could be a valuable source of information for multiple purposes, including quality management of patient care plus its

documentation, and thus wanted to tap the supposed potential within these data.

This afforded an opportunity to apply the raw framework for the development of secondary use solutions in a definite problem domain which could serve as a model-domain for routine clinical data (paradigmatic case selection [24]). The nursing field was indeed very interesting as a problem domain, as nursing data is generated in large volumes during all phases of the patient stay, it is multifaceted and process-based, and it can contain structured, semi-structured, and completely unstructured elements as well as long sections of free text. In addition, there exist a lot of non-uniform or institution-specific documentation and coding standards in the nursing domain.

Besides developing a working and quite useful reuse solution for the nursing management board, our superordinate goal in this case study was to learn from the planning and implementation in order to improve and refine the raw SPIRIT framework. We also wanted to elaborate a preferably detailed best-practice procedure model based on the insights gained in this study. In the following, we give a short overview on the achievements from Case Study I. More details concerning this case study and the resulting nursing intelligence system can be found in [25].

The realized solution included detailed concepts for data extraction and transformation, for data integration and storage, and for data use and analysis. These concepts were implemented based on open source software tools following an iterative development approach. The resulting solution, a nursing intelligence system consisting of a nursing data mart for data integration and a web-based analysis platform, are in routine use at tirol kliniken. Nursing data from all inpatient stays since 11/2009 have been imported on a monthly basis. The nursing data mart, for example, contains more than one-and-a-half million nursing diagnoses and care paths. Over 27 million single data elements from nursing care planning and evaluation are available at the moment. Around 70 adaptable reports were designed to address the defined analysis questions of the tirol kliniken nursing management board. In addition to

the reporting functionality, the nursing intelligence system allows more sophisticated analysis techniques, such as online analytical processing (OLAP) and custom queries, to be implemented in order to address individual analysis questions or to extract data subsets for further data mining operations.

Based on the experiences and lessons learned during the iterations of the developmental cycles, the raw SPIRIT framework was further refined. The skeleton of the framework, a detailed procedure model listing all necessary steps and related tasks, was also elaborated during these iterations. In order to avoid overfitting the framework and procedure model to the specific conditions in the case study, each developmental cycle was followed by a phase of generalization. In these phases, all developments underwent a notional re-engineering run in which boundary conditions were altered hypothetically (e.g. differing healthcare institutions, underlying documentation systems, classification systems and terminologies, additional or missing data elements, altered clinical context and processes, data types, differing questions for analyses).

3.3 Case Study II (Multimodal Intensive Care and Neuromonitoring Data): Shaping SPIRIT's Body

In a second case study, the refined SPIRIT framework was used to develop a comprehensive routine data reuse concept for the Neurological Intensive Care Unit (ICU) at Innsbruck Medical University, Department of Neurology. The ICU has 18 beds, 12 for intensive care (IC) plus 6 for intermediate care (IMC), and treats approximately 800 (500 IC, 300 IMC) patients per year. The average length of stay is 6.7 (IC) and 6.3 (IMC) days, respectively (key figures from 2013). During these stays, an immense mass of multimodal high-resolution patient monitoring data is produced by a multitude of medical devices in addition to the „normal" routine care data (patient characteristics, diagnoses, medication information, nursing care data, treatment data, etc.).

The intention of this second case study was to apply SPIRIT in the field and to critically reflect its applicability in a com-

pletely different clinical context. Based on the lessons learned from this second case study, SPIRIT was further refined. As a by-product due to an accurate project controlling within this case study, we were able to determine the effort for the development of such a routine data reuse concept developed with SPIRIT.

The ICU routine data reuse concept comprised a compilation of generic question classes, including the explicit topical questions of the ICU's medical management board and a 3LGM^2 model of the ICU sub-information system. 3LGM^2 stands for *three-layer graph-based meta model*. It is a useful approach to model comprehensive clinical information systems. The approach distinguishes three different, interconnected layers of information systems: i) the domain layer (*"organizational perspective"*), which describes the enterprise functions (e.g. clinical tasks) together with relevant information objects; ii) the logical tool layer (*"software perspective"*), which describes the application components of the information system with data communication (e.g. interfaces, communication links and standards, protocols) and data storage components (e.g. database management systems); and iii) the physical tool layer (*"hardware perspective"*), which describes the physical data processing components (e.g. medical devices, client PCs, networks, servers) [26, 27].

The ICU routine data reuse concept further comprised a complete list of available data elements plus available metadata (e.g. origin, data type, scale and range, unit), which was linked with the 3LGM^2 model, a concept for data extraction, transformation and loading (ETL), and a proposal for a multidimensional data model with implementation recommendations. The total personnel expenditure for the elaboration of the concept was 460 person-hours (including project management, organization, and report generation). A pilot data mart with aggregated multimodal ICU data was established to test the feasibility of the approach. Several findings resulting from this pilot data mart could already be published (e.g. [28–32]).

In this second case study, we essentially learned more about the usefulness for further planning phases of the reuse solution of such a detailed model of the investigated potential source information systems. With the 3LGM^2 model and the linked list of available data elements, we could easily identify whether certain information objects needed to answer particular analysis questions were available in our source systems. We could further determine the optimal place to tap these data elements (e.g. to minimize the number of necessary data interfaces if data elements were stored in multiple locations or in different formats, or if different communication standards were used). Furthermore, we learned about the importance of a strategic alignment of the reuse solution's goals with the overall business goals of the healthcare (sub-)organization

Based on all lessons learned during this second case study, the SPIRIT framework was finally refined and consolidated. In the following, we want to present SPIRIT in detail.

4. Results – The SPIRIT Framework

4.1 Overview of SPIRIT

The SPIRIT framework can be used to systematically plan and design solutions for the intelligent reuse of clinical routine data. SPIRIT aims at tailoring such secondary use solutions to specific needs of single departments without losing sight of the whole institution. It provides a conceptual best-practice framework consisting of three main blocks – ①*Vision & Strategy*, ②*Analysis & Specification*, and ③*Feasibility & Implementation* – which are executed sequentially and can be iterated if required during the lifecycle of the secondary use solution. Each of these blocks again contains a procedure model with different modules, which in turn contain sets of required tasks. The execution order of these tasks is not strictly fixed. It may vary as a function of the present clinical context with its special conditions or local requirements. Some tasks should be executed in parallel (see below). ▶Figure 1 gives an overview of the conceptual framework and the different modules of the SPIRIT procedure model. For a more detailed descrip-

tion of the single modules and tasks, see below. Additional details can be found in [33] (available in German).

4.2 ① Vision & Strategy

The first block to be executed is to define a secondary use strategy comprising vision and goals of the intended routine data reuse on a strategic level. In particular, a holistic perspective of the intended routine data reuse is very important because different stakeholders and future users of a clinical routine data reuse solution may have different requirements and may pursue different individual goals. Often, single departments or their leaders pioneer such developments and other departments later follow their example. Secondary use solutions thus usually evolve successively, making it crucial to align different individual goals strategically with the vision and development plan for the entire healthcare organization already at the beginning of the planning phase.

The customer or purchaser of the solution, together with members of the management board of the healthcare organization, should answer the following principal questions:

- Does a vision for the healthcare organization exist and what is it?
- What are the strategic goals and business objectives for the healthcare organization?
- Which strategy does the management of the healthcare organization pursue?
- What are the main objectives for a reuse of clinical routine data?
- Who and where are the definite and potential users?
- How can the secondary use solution be strategically aligned with the strategic goals and business objectives of the healthcare organization?
- What should be the level of availability of a secondary use solution (e.g. enterprise- or department-wide, only for specific user groups, etc.)?
- How can analytical findings, derived from routine data, support vision, strategy, and goals?
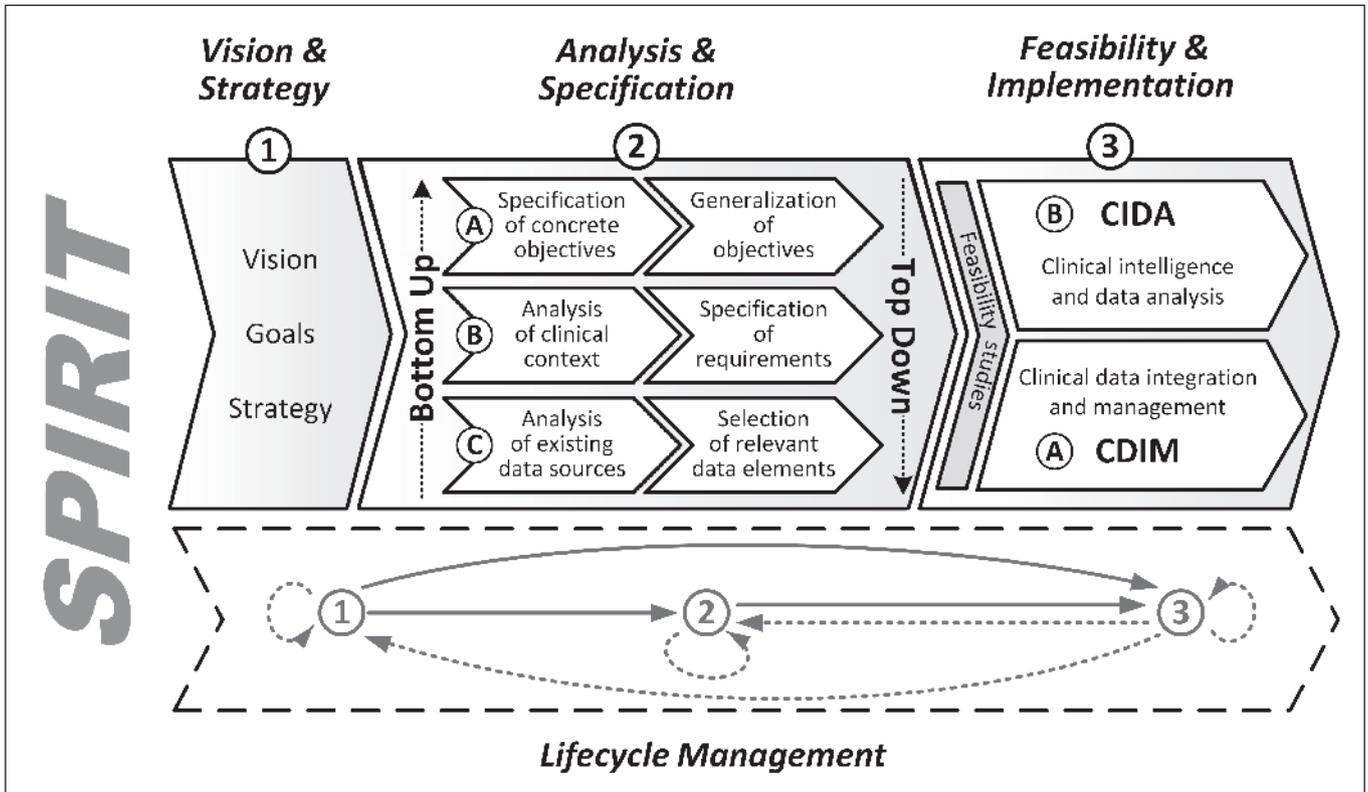- What are the immediate and long-term objectives of the customer or purchaser

Figure 1 Overview of the conceptual framework with three main blocks (①–③). These blocks contain the different modules of the SPIRIT procedure model. The dashed rimmed block at the bottom depicts the execution order and flow of actions during the implementation phase (solid arrows) and life cycle (dotted arrows) of the implemented clinical routine data reuse solution.

and known future users of the secondary use solution?

Suitable methods for obtaining the answers to these questions are, for example, expert interviews or focus groups. Additionally, workshops with all relevant stakeholders should be conducted to reach a consensus on these questions. ▶Figure 2 presents typical objectives for routine data reuse. It can be used in such workshops as look-up pattern to facilitate the definition of strategic goals for the reuse solution. It was developed based on the inductive classification of areas of application for clinical routine data reuse that were found in the relevant publications from the literature overview and based on the lessons learned in the two case studies.

This first SPIRIT block is of significant importance for all subsequent activities. The secondary use strategy forms the starting basis for a set of comprehensive analyses to be conducted in the second SPIRIT block. The results of this first block also

have implications for both partitions in the third block, for the principal architecture of the clinical data integration and management layer (CDIM), and for the design of the clinical intelligence and data analysis layer (CIDA) of the final solution (e.g. enterprise-wide or local focus only, independent or dependent data mart architecture, real or virtual resulting clinical data warehouse, organization of the analysis platform, supported analysis techniques, authorization concept, rollout plan, etc.).

4.3 ② Analysis & Specification

The success of any secondary use solution significantly depends on the perceived usefulness of information gained out of it (i.e. *"Big Information"*) among the users. Consequently, the central block within SPIRIT aims at specifying the analysis questions and requirements of the different stakeholders as well as thoroughly analysing prerequisites and boundary conditions, such as available data sources, concrete

data elements, and clinical workflows. For this, in-depth analyses from the future users' and strategic perspective (TOP DOWN analysis track–Ⓐ) and from the system and data level (BOTTOM UP analysis track–Ⓒ) have to be conducted. A mandatory, comprehensive analysis of the clinical context, in which the data is produced and "Big Information" will be used, helps to complement these two contrarian analysis tracks (IN BETWEEN analysis track–Ⓑ).

All three tracks comprise two phases that should be executed sequentially. The three analysis tracks themselves should be executed concurrently starting with the TOP DOWN track. Insights gained in each track should be reflected in the others. Therefore, it is important that the leaders responsible for the different analysis tracks coordinate their efforts and collaborate closely.

4.3.1 TOP DOWN Analysis Track

The TOP DOWN approach starts with the specification of future users' concrete objectives. The goal of the first phase is to select candidates for all conceivable kinds of analysis questions that future users want answered by analysing routine data. For this, expert interviews, focus groups, and creativity workshops should again be conducted. In addition, literature reviews can be conducted to collect relevant questions.

In a second generalization phase, the secondary use strategy as defined in the previous SPIRIT block should be used as a starting point. The different concrete analysis questions, as defined by the future users, must be clustered into coherent groups coinciding with the reuse strategy.

This should be done within the project team during a separate workshop and will help to design CDIM and CIDA as flexibly as possible in the third block of SPIRIT.

4.3.2 BOTTOM UP Analysis Track

The BOTTOM UP analysis track starts with a detailed systems analysis in which all of the existing data sources are registered and described systematically. For this, in the relevant clinical sub-information systems, all application systems used to create, record, document, process, or store clinical routine data have to be listed. In order to identify all relevant systems, results from the first phase of the IN BE-

TWEEN analysis track (clinical context analysis, see below) will be helpful.

The resulting list should comprise type, purpose, name, and vendor of the application systems plus a description, as detailed as possible, of all processed and stored data elements or information objects including available metadata (e.g. data type, scale and range, unit, origin, storage type, and location). If needed, the software and hardware components of these application systems should also be listed. Furthermore, the available data communication interfaces of the application systems and communication standards used should be analysed and documented, as this information can be valuable for the design of the data extraction strategy.
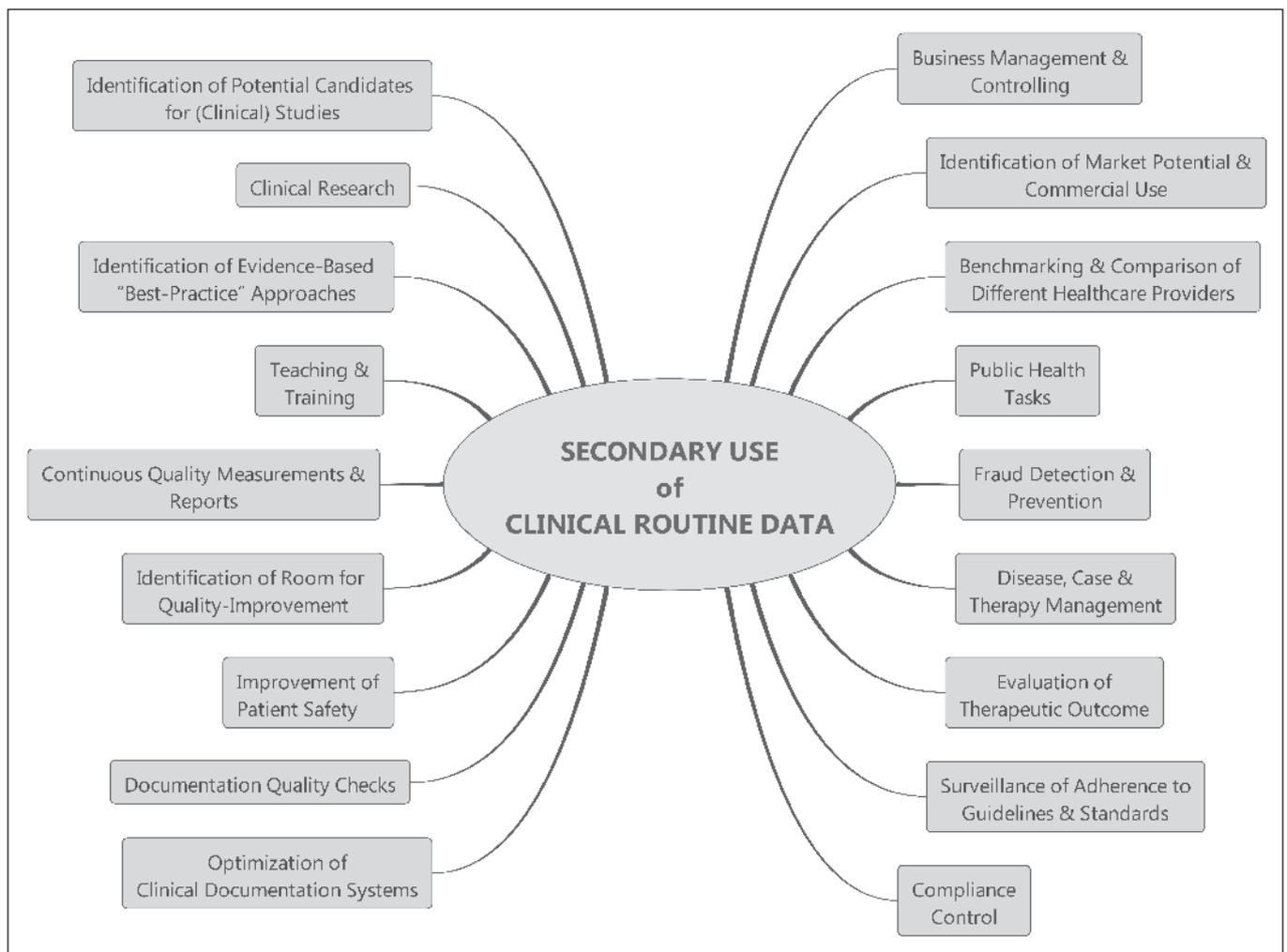

Figure 2   Possible classification of objectives for clinical routine data reuse (developed during Case Studies I and II, based on results from the literature overview)

In order to receive a better overview of the application systems, of the flow of data elements through these systems, and of their storage locations, a more formal representation of the results of this system analysis may be required. This may be the case, in particular, when dealing with complex information system architectures with large numbers of application systems and data elements or multiple storage locations for certain data elements (e.g. as in Case Study II – multimodal data reuse concept for a neurological intensive care unit). A modelling technique that is suited for this purpose is 3LGM^2, the three-layer graph-based meta model [26, 27].

Having such a model is also very helpful in the second phase of the BOTTOM UP approach when the relevant data elements to be integrated in the secondary use solution are selected in preparation for the design of a data extraction strategy in the next SPIRIT block. The collection of analysis questions resulting from the first phase of the TOP DOWN analysis track is needed as a prerequisite for this step. Information on the clinical context (e.g. workflows and clinical processes, involved professional groups, etc.) can also be of value for this phase.

4.3.3 IN BETWEEN Analysis Track

The analysis of the clinical context offers valuable clues to the circumstances in which the clinical routine data is produced. The secondary use strategy, including the answers to the questions from the first block, builds the starting point for the context analysis. First, key facts concerning the healthcare organization (e.g. type and form of organization, sponsorship, overview of key enterprise functions and service offers, central diagnostic and therapeutic procedures, fields of specialization, organizational structure, plan and management, budget, personnel structure, staff appointment schemes, etc.) and key figures (e.g. available beds and level of capacity, number and length of patient stays, outpatient visits, catchment area, treatments, diagnoses and DRGs, orders, etc.) have to be gathered from the institution's controlling department or from annual reports. The results allow an estimation of the

amount of available routine data and reuse possibilities.

In a next step, the relevant diagnostic, therapeutic, and supportive processes must be carefully examined. Special attention must be given to the clinical documentation. The following aspects should be analysed:

- Which diagnostic, therapeutic, and supportive processes are executed in the institution?
- How can these processes be characterized?
- Which data are documented in which process steps?
- Who carries out the documentation?
- Where, when, and in which form (e.g. degree of structuring) does the documentation take place?
- Which documentation standards are used?
- Which tools (e.g. documentation systems, electronic forms) are used?
- What is the primary purpose of usage of the documented data?

This should be done through observations during site visits and by interviewing healthcare professionals in the field. Forms, input masks, or documented data can also be analysed. A very practical method which should also be applied is to run through fictitious cases of different test patients and to document these cases using the existing documentation systems. Data generated in this way can then be used to design and to test the data extraction, transformation and loading (ETL) routines. However, this should be done only in addition to the observations and interviews in order to avoid receiving only an idealized picture of the real situation.

In the second phase of the IN BETWEEN track, the results of the TOP DOWN and BOTTOM UP tracks are used to compile all functional and non-functional requirements regarding the intended reuse solution. After finalizing the requirements specification, the next SPIRIT block can start.

4.4 ③ Feasibility & Implementation

As noted above, a clinical routine data reuse solution can never be available out of the box. It has to be tailored to the specific requirements and context as analysed and defined in the previous SPIRIT block. The third SPIRIT block, *"Feasibility & Implementation"*, is thus dedicated to the design and implementation of the routine data reuse system, which consists of two layers: CDIM (Ⓐ – clinical data integration and data management) and CIDA (Ⓑ – clinical intelligence and data analysis). It should be noted that implementation must be understood as the adaptation and assembly of different components to a best-of-breed solution. A large variety of open source and commercial data warehousing and data integration solutions are available on the market. However, these are often not mature enough to be implemented without modification, or they do not meet all defined requirements of the planned reuse solution, so that a complete ready-to-use solution may not be available. It may thus be necessary for single components to be developed anew if no adequate solution for a specific purpose can be found.

4.4.1 Preceding Feasibility Studies

Comprehensive feasibility studies are conducted at the start in order to deliver essential information on framework conditions relevant for the design and the subsequent implementation phases (e.g. estimation of accumulating data volumes, determination of needed resources, acquisition of pending permissions and final clarification of data security and privacy issues, establishment of contacts to all data suppliers, selection of a software development model, setup of programming and testing environments, testing and selecting of hardware and software components, delineation of the data extraction strategy, the ETL routines, a data model, an analysis framework, etc.).

4.4.2 CDIM – Clinical Data Integration and Data Management

To build the fundamental layer of a clinical routine data reuse system, a data extraction

strategy must first be developed. This strategy defines how the raw routine data elements that had been classified as relevant in the BOTTOM UP track in the second block are extracted from the existing source systems. The strategy also has to define rules for conflict cases (e.g. in the case of discrepancies among redundant data elements, such as differing date and time data or conflicting primary diagnoses in different systems). It must define the optimal place to tap the data sources in order to guarantee optimal data quality along with limiting efforts for interface development. The strategy should also define the extraction periodicity and latency.

After defining the extraction strategy, the single ETL routines must be defined and implemented. A further step to be carried out hand in hand with the design of the data transformation is the definition of a multidimensional data model. It allows the integration of the extracted data to answer all of the analysis questions which were requested from the future users in the TOP DOWN track in the second block. Moreover, the data model must be easily extendible and should allow a flexible integration of new data elements or dimensions. With regard to the holistic perspective of a clinical routine data reuse system, the data model must not be fitted too tightly to the situation in one specific department. It should contain a hard core, a kind of minimal data set, which is valid for other departments as well as in the overall organization.

Then, depending on the development model chosen, one or multiple implementation and testing phases are passed through until a final version of the CDIM layer is available.

4.4.3 CIDA – Clinical Intelligence and Data Analysis

The end users' interaction with the final clinical routine data reuse solution will primarily happen through the components of the CIDA layer. In an optimal case, an end user will experience one comprehensive „Clinical Intelligence System". The overall user experience of this system will mostly depend on the design, functionality, and perceived usefulness of the single components in the CIDA layer.

The design and implementation of the data analysis layer goes hand in hand with the design of the CDIM layer. The first task in the design phase of this layer is to define the principal architecture of a data analysis framework (e.g. stand-alone application or multitier architecture with thin or fat client applications, analysis tools, etc.). An access control and authorization concept must be developed. The required analysis techniques (e.g. querying, simple reporting, ad-hoc reporting, OLAP, data mining and machine learning, semantic mining, etc.) and the appropriate tools then have to be selected. In addition, the presentation of analysis results (e.g. charts, tables, dashboards, etc.) should be designed.

The realization of sophisticated data analyses is the primary purpose of any secondary use application. It is therefore crucial to create a powerful data analysis workbench that meets all defined analysis requirements and is able to answer all analysis questions as defined in the previous block. Special attention should also be paid to usability and performance issues of the analysis components, as these have a strong influence on the user experience. Comprehensive tests must be performed during and subsequent to the implementation phases. Finally, in line with the roll-out of the completed system, intensive training sessions with the end-users should be held.

4.5 Lifecycle Management

As depicted previously, clinical information systems evolve constantly and change over time. This fact already needs to be taken into account for every clinical routine data reuse system. Therefore, the last block of SPIRIT is dedicated to lifecycle management. In every case of significant change among the clinical information system, the clinical context or requirements concerning analyses, whole SPIRIT blocks, or single modules within these blocks can be iterated to adapt the clinical routine data reuse system to the new situation.

5. Discussion
5.1 SPIRIT's Assets

SPIRIT is a conceptual best-practice framework including a procedure model that facilitates putting into practice the idea of secondary use within the hospital by systematically planning the intelligent reuse of clinical routine data for multiple purposes.

Such comprehensive clinical routine data reuse solutions are usually too complex to be realized in a "big bang" approach. Of course, the SPIRIT framework would be suitable to plan such a "sweeping blow" project, but it seems more feasible to adopt the "divide and conquer" idea, i.e. to start small and to implement special solutions for particular application purposes or locations while keeping the vision of Big Clinical Data in mind. In more specific terms, this means that the small solutions must be designed in a flexible and scalable manner, and that all the single solutions must share a common core data set with shared key indices (e.g. unique patient IDs, case IDs, clinical process IDs, etc.). Such a shared core minimum data set allows diverse data linkage and further data aggregation steps and enables the subsequent creation of a physically or virtually integrated enterprise-wide clinical routine data reuse repository and a comprehensive clinical intelligence system.

SPIRIT allows for such a stepwise line of action and starts with the determination of a secondary use strategy for the whole organization. After defining the reuse goals on a strategic level, two-phase, comprehensive, site- and purpose-directed analyses are conducted from two viewpoints: from a systems level and the perspective of available data (BOTTOM UP) and from a future users' point of view (TOP DOWN) to precisely define the requirements regarding a clinical routine data reuse solution. The mandatory clinical context analysis (IN BETWEEN) helps to join and refine the requirements resulting from these two polar perspectives. During the following design and implementation phases, an optimal, two-layer, best-of-breed solution that matches the respective application purposes and/or locations can be realized.

As shown in the case studies, SPIRIT is applicable in practice and the solutions for

reusing clinical routine data, as designed following the SPIRIT framework, achieve the primary goal of data reuse. That is the creation of new insights and knowledge from existing data. The nursing intelligence system designed and implemented in Case Study I is in routine use for quality management of the nursing process and nursing documentation at tirol kliniken. The data mart with multimodal ICU data, as designed and piloted in Case Study II, yielded new insights into intensive care [34] and patient safety, which could be documented in several publications.

5.2 Limitations

SPIRIT was developed as a best-practice approach based on a non-systematic literature overview. It was refined according to lessons learned in two separate case studies in completely different clinical settings. On the one hand, this fact can be seen as a clear strength, as the selection of two paradigmatic cases ("having strategic importance in relation to the general problem") with a maximum of variability [24] ensures SPIRIT's practical applicability and limits the risk of an overfitting to certain boundary conditions in a single case study. On the other hand, no formal evaluation or validation of SPIRIT was performed. Therefore, evidence for the framework's universal validity is still not guaranteed. However, in addition to a well-considered selection of the cases, the methodical design of the two case studies aimed at yielding a largely open and generally applicable best-practice framework and procedure model.

The SPIRIT framework was developed in a hospital setting, the intended primary usage site. SPIRIT is also applicable in other healthcare settings (e.g. outpatient care, long-term care facilities, etc.) but this may require adoptions of parts of the framework to the specific setting. Especially for the design of trans-institutional and integrated, multicentre-fed routine data reuse solutions, additional considerations and tasks may be required (e.g. creation and administration of master patient indices [35], definition of interfaces and terminology servers for appropriate mappings, realization of semantic interoper-

ability [36], etc.). Moreover, we may expect that data volumes will continue to grow exponentially and that, in the future, the approach of aggregating clinical routine data in centralized data warehouses will be discarded in favour of distributing the computation. Warehouses may become virtual and be distributed over different storage locations. Secondary use strategies will have to take this into account.

The description of the single tasks within the SPIRIT procedure model is rather general. An appropriate methodology for the individual modules must be elaborated for the particular cases. SPIRIT is also independent of details with regard to design and implementation, as these depend on the individual cases. Last but not least, specifications concerning project management and controlling are also not further specified. A variety of literature is available for this purpose, e.g. [37–39].

5.3 SPIRIT in Relation to Other Approaches

As depicted above, the principal purpose of any clinical routine data reuse solution should be the generation of new knowledge from existing clinical routine data. A large variety of publications reporting successful discoveries of new knowledge can be found [40] and there continued hype and expectations exist regarding secondary use and big data in healthcare [41]. This implies that a large amount of clinical routine data resources exists and is being used for diverse purposes. For the technical issues, different solutions are available and questions on aspects of how to store and process such large amounts of data can be answered [42]. Complete architectures for sophisticated analytical platforms are also described, e.g. [43, 44].

However, it is very interesting that, on the other hand, answers to questions of how to plan and establish such resources in a systematic way are difficult to find. This is nothing new. Already ten years ago, Sen and Sinha stated that data warehousing methodologies were rapidly evolving, but that the field of data warehousing in general was not very mature [45]. They compared 15 data warehousing methodologies and concluded that no method had by then

achieved the status of a recognized standard. Methodologies and procedure models that were developed or promoted by hardware manufacturers focused on technological aspects, whereas software vendors fostered software-driven approaches. Samaha and Croll came to a similar conclusion for the health informatics domain as they experimented with these methodologies to test if they were suitable for establishing clinical data warehouses [46]. They underlined the importance of analysing the business context and the underlying clinical processes, which was missing in some methodologies. In SPIRIT, the context analysis is a central task. An explicit lifecycle management is also included in the SPIRIT approach. This was missing in the majority of the other investigated approaches as stated in [46]. In an evaluation of different data warehousing tools, Do et. al. criticized insufficient or missing metadata management possibilities [47]. Especially in the constantly evolving clinical domain, with changing knowledge, procedures, information systems, and data, an appropriate metadata management capable of dealing with a variety of so-called slowly changing dimensions [48] is of high importance. The BOTTOM UP analysis in the second SPIRIT block determines all relevant meta-information on the available data objects and lays the foundations for the design of an appropriate metadata management to be implemented in the CDIM layer.

For the data analyses, a plethora of methods and tools are also available. The KDD process (knowledge discovery in databases) as proposed by Fayyad et. al. in 1996 [49], for example, can be seen as ancestor of all systematic approaches aiming at the targeted generation of new knowledge from existing data. KDD is commonly used and requires five steps: i) data selection, ii) pre-processing, iii) transformation, iv) data mining, and v) interpretation/evaluation to be performed for each analysis. OLAP (online analytical processing), as proposed in 1993 by Codd [50], is another widely used approach for the analysis of large databases. OLAP tools support interactive data analysis operations such as data breakdowns and summaries along defined dimensions. SPIRIT does not dictate the

use of specific analysis approaches or methods. It aims at establishing a powerful data intelligence and analysis workbench which is exactly tailored to the needs of the users but can be extended or adapted to changing requirements (CIDA layer).

5.4 Open Questions and Outlook

Neither legal and regulatory aspects nor data privacy and security matters are addressed in the SPIRIT framework. Such questions must be resolved based on the local circumstances and in compliance with the applicable law.

The different analysis tracks require much time and resources. The cost-benefit ratio of the SPIRIT approach has not yet been evaluated and it is not clear whether a systematic planning of routine data reuse, as proposed, brings measurable benefits or better results than a non-systematic approach, which simply accumulates and uses the available data.

Proof of the framework's feasibility for the creation of trans-institutional and multicentric routine data repositories also remains open. At the moment, SPIRIT is used to develop a concept for a long-term (> 20 years), multicentric, multimodal clinical and routine data registry platform for the prospective evaluation of the radiotherapeutic dose distribution as the cause for induction of secondary malignoma [51]. In our group, SPIRIT will also be used to expand the nursing routine data reuse solution developed in Case Study I to test the feasibility of transforming routine nursing data into a nursing minimum data set [52] and to establish a patient safety intelligence system.

6. Conclusion

The SPIRIT framework can be used to systematically plan the intelligent reuse of clinical routine data for multiple purposes. It helps to establish a strategic data reuse management plan which can be embedded into a strategic information management plan of the whole organization. Based on this strategic alignment, SPIRIT allows a stepwise setup and evolution of holistic clinical routine data reuse solutions.

References

1. Berner ES, Moss J. Informatics challenges for the impending patient information explosion. J Am Med Inform Assoc 2005; 12 (6): 614–617.
2. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE, Expert P. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007; 14 (1): 1–9.
3. Wade TD. Refining gold from existing data. Curr Opin Allergy Clin Immunol 2014; 14 (3): 181–185.
4. Aller RD. The clinical laboratory data warehouse. An overlooked diamond mine. Am J Clin Pathol 2003; 120 (6): 817–819.
5. Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, Leese A, Richardson C, Mantas J, Murray P, De Moor G. Trustworthy reuse of health data: a transnational perspective. Int J Med Inform 2013; 82 (1): 1–9.
6. Bréant C, Thurler G, Borst F, Geissbuhler A. Design of a Multi Dimensional Database for the Archimed DataWarehouse. Studies in Health Technology and Informatics 2005; 116: 169–174.
7. Canadian Institutes of Health Research. Secondary Use of Personal Information in Health Research: Case Studies. Ottawa: Public Works and Government Services Canada; 2002. p 150.
8. 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 (2): 131–135.
9. Dorda W, Gall W, Duftschmid G. Clinical data retrieval: 25 years of temporal query management at the University of Vienna Medical School. Methods Inf Med 2002; 41 (2): 89–97.
10. Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU. Unlocking Data for Clinical Research – The German i2b2 Experience. Appl Clin Inform 2011; 2 (1): 116–127.
11. Junttila K, Meretoja R, Seppälä A, Tolppanen E-M, Ala-Nikkola T, Silvennoinen L. Data warehouse approach to nursing management. Journal of Nursing Management: Blackwell Publishing Limited; 2007. pp 155–161.
12. Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. AMIA Annu Symp Proc 2012; 2012: 189–198.
13. de Mul M, Alons P, van der Velde P, Konings I, Bakker J, Hazelzet J. Development of a clinical data warehouse from an intensive care clinical information system. Comput Methods Programs Biomed 2012; 105 (1): 22–30.
14. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. AMIA Jt Summits Transl Sci Proc; 2013. pp 6–8.
15. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, Gainer V, Berkowicz D, Glaser JP, Kohane I, Chueh HC. Architecture of the opensource clinical research chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc 2007. pp 548–552.
16. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. JAMA 2014; 312 (15): 1513–1514.
17. Boyce RD, Ryan PB, Noren GN, Schuemie MJ, Reich C, Duke J, Tatonetti NP, Trifiro G, Harpaz R, Overhage JM, Hartzema AG, Khayter M, Voss EA, Lambert CG, Huser V, Dumontier M. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Saf 2014; 37 (8): 557–567.
18. Peters SG, Buntrock JD. Big data and the electronic health record. J Ambul Care Manage 2014; 37 (3): 206–210.
19. Bellazzi R. Big Data and Biomedical Informatics: A Challenging Opportunity. Yearb Med Inform 2014; 9 (1).
20. Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 2009; 48 (1): 38–44.
21. Inmon B. Data warehousing in a healthcare environment. Pittsburgh (PA): The Data Administration Newsletter, LLC (TDAN.com); 2007 [cited 01.10.2014]. Available from: http://www.tdan.com/view-articles/4584/.
22. Leiner F, Haux R. Systematic planning of clinical documentation. Methods Inf Med 1996; 35 (1): 25–34.
23. Coiera E. Guide to health informatics. London; New York, NY: Arnold; Distributed in the USA by Oxford University Press; 2003.
24. Flyvbjerg B. Five Misunderstandings About Case-Study Research. Qualitative Inquiry 2006; 12 (2): 219–245.
25. Hackl WO, Rauchegger F, Ammenwerth E. A Nursing Intelligence System to Support Secondary Use of Nursing Routine Data. Appl Clin Inform 2015; 6 (2): 418–428.
26. Winter A, Brigl B, Funkat G, Haber A, Heller O, Wendt T. 3LGM(2)-Modelling to Support Management of Health Information Systems. Stud Health Technol Inform 2005; 116: 491–496.

27. Winter A, Haux R. A three-level graph-based model for the management of hospital information systems. Methods Inf Med 1995; 34 (4): 378–396.

28. Antunes AP, Schiefecker AJ, Beer R, Pfausler B, Sohm F, Fischer M, Dietmann A, Lackner P, Hackl WO, Ndayisaba JP, Thome C, Schmutzhard E, Helbok R. Higher brain extracellular potassium is associated with brain metabolic distress and poor outcome after aneurysmal subarachnoid hemorrhage. Crit Care 2014; 18 (3): R119.

29. Helbok R, Schiefecker A, Delazer M, Beer R, Bodner T, Pfausler B, Benke T, Lackner P, Fischer M, Sohm F, Hackl W, Stover JF, Thome C, Humpel C, Schmutzhard E. Cerebral tau is elevated after aneurysmal subarachnoid haemorrhage and associated with brain metabolic distress and poor functional and cognitive long-term outcome. J Neurol Neurosurg Psychiatry 2015; 86 (1): 79–86.

30. Schiefecker AJ, Pfausler B, Beer R, Sohm F, Sabo J, Knauseder V, Fischer M, Dietmann A, Hackl WO, Thome C, Schmutzhard E, Helbok R. Parenteral diclofenac infusion significantly decreases brain-tissue oxygen tension in patients with poor-grade aneurysmal subarachnoid hemorrhage. Crit Care 2013; 17 (3): R88.

31. Helbok R, Schiefecker A, Beer R, Dietmann A, Antunes AP, Sohm F, Fischer M, Hackl WO, Rhomberg P, Lackner P, Pfausler B, Thome C, Humpel C, Schmutzhard E. Early brain injury after aneurysmal subarachnoid hemorrhage: a multimodal neuromonitoring study. Critical Care 2015; 19 (1): 75.

32. Kofler M, Schiefecker A, Ferger B, Beer R, Sohm F, Broessner G, Hackl W, Rhomberg P, Lackner P, Pfausler B, Thome C, Schmutzhard E, Helbok R. Cerebral Taurine Levels are Associated with Brain Edema and Delayed Cerebral Infarction in Patients with Aneurysmal Subarachnoid Hemorrhage. Neurocrit Care 2015;23 (3): 321–329.

33. Hackl WO. Erschließung und Sekundärnutzung von Routinedaten aus der klinischen und pflege-

rischen Prozessdokumentation: Ein Rahmenkonzept, Vorgehensmodell und Leitfaden [Dissertation]. Hall in Tirol: UMIT; 2013.

34. Wolf S. Take care when taking care of fever after aneurysmal subarachnoid hemorrhage. Crit Care 2013; 17 (4): 160.

35. Dooling JA, Downing K. Ensuring data integrity through a clean master patient index. J AHIMA 2014; 85 (3): 46–47.

36. Zunner C, Ganslandt T, Prokosch HU, Burkle T. A reference architecture for semantic interoperability and its practical application. Stud Health Technol Inform 2014; 198: 40–46.

37. Houston SM, Bove LA. Project management for healthcare informatics. New York; London: Springer; 2010.

38. Coplan S, Masuda D. Project management for healthcare information technology. New York: McGraw Hill; 2011.

39. Ammenwerth E, Haux R, Knaup-Gregori P, Winter A. IT-Projektmanagement im Gesundheitswesen. Lehrbuch und Projektleitfaden; taktisches Management von Informationssystemen; mit 84 Tabellen. Stuttgart: Schattauer; 2015.

40. Ganslandt T, Hackl WO. Findings from the Clinical Information Systems Perspective. Yearb Med Inform 2014; 10 (1): 90–94.

41. Wang W, Krishnan E. Big Data and Clinicians: A Review on the State of the Science. JMIR Medical Informatics 2014; 2 (1): e1.

42. Peek N, Holmes JH, Sun J. Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics. Yearb Med Inform 2014; 9 (1): 42–47.

43. Post AR, Kurc T, Cholleti S, Gao J, Lin X, Bornstein W, Cantrell D, Levine D, Hohmann S, Saltz JH. The Analytic Information Warehouse (AIW): a platform for analytics using electronic health record data. J Biomed Inform 2013; 46 (3): 410–424.

44. Mandl KD, Kohane IS, McFadden D, Weber GM, Natter M, Mandel J, Schneeweiss S, Weiler S, Klann JG, Bickel J, Adams WG, Ge Y, Zhou X, Perkins J, Marsolo K, Bernstam E, Showalter J, Quar-

shie A, Ofili E, Hripcsak G, Murphy SN. Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): architecture. J Am Med Inform Assoc 2014; 21 (4): 615–620.

45. Sen A, Sinha AP. A Comparison of Data Warehousing Methodologies. Communications of the ACM 2005; 48 (3): 79–84.

46. Samaha TR, Croll PR. A Data Warehouse Architecture for Clinical Data Warehousing. In: Roddick F, Warren JR, editors. Proceedings Australasian Workshop on Health Knowledge Management and Discovery (HKMD 2007) CRPIT; Ballarat, Victoria: Australian Computer Society; 2007. pp 227–232.

47. Do HH, Stöhr T, Rahm E, Müller R, Dern G. Evaluierung von Data Warehouse-Werkzeugen. In: Jung R, Winter R, editors. Data Warehousing 2000: Methoden, Anwendungen, Strategien [Proceedings der Tagung "Data Warehousing 2000" am 14. und 15. November 2000 in Friedrichshafen]. Heidelberg: Physica-Verl.; 2000. pp 44–57.

48. Leonard A, Masson M, Mitchell T, Moss J, Ufford M. Slowly Changing Dimensions. SQL Server 2012 Integration Services Design Patterns: Apress; 2012. pp 261–272.

49. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. AI MAGAZINE 1996; 17 (3): 37–54.

50. Codd EF, Codd SB, Salley CT. Providing OLAP (On-line analytical processing) to user-analysts: an IT mandate. Ann Arbor, Mich.: Codd & Associates; 1993.

51. COMET Center Oncotyrol. SEMPER: SEcondary Malignoma – Prospective Evaluation of the Radiotherapeutic dose distribution as the cause for induction (2.4.5). 2014. Available from: http://www.oncotyrol.at/forschung/forschungsprojekte/area-4-hta-and-bioinformatics/245/.

52. Ranegger R, Hackl WO, Ammenwerth E. A proposal for an Austrian Nursing Minimum Data Set (NMDS): a Delphi study. Appl Clin Inform 2014; 5 (2): 538–547.