1. Introduction

Developments in the field of computer-based patient records (CPRs) have been ongoing for nearly 3 decades. The introduction of CPRs in primary care has been especially successful in the UK and the Netherlands [1]. CPR development for specialized care has been most successful in the US. Pioneer endeavors resulted in the well-known systems RMIS, COSTAR, and TMR [2-5]. These systems initially focused on the administrative and financial aspects of health care, the electronic patient record being a later development. Experiences with these systems in routine care show that physicians are much quicker to use computers for consultation than for data recording, which is mainly done via written encounter forms. The collection of structured data at the point of care, however, offers many benefits for research, decision support, and quality assessment [6-8]. Despite these benefits, many new applications that support structured data entry (SDE) have not been well received in routine practice. A part from tight time schedules, clinicians need to invest before they can harvest the benefits: the more data is available in a structured format, the greater the benefits.

Systems that support SDE can suffer from lack of flexibility, scope, or detail. Fixed forms impose limits on the data items covered and are often a compromise between the needs of different users. Many forms not only make a system cumbersome and difficult to maintain, but may also introduce redundancy and inconsistency at the data level. Menu-driven applications are more flexible, but often time-consuming or limited in domain and expressiveness [9-12].

Despite the introduction of computers in health care, patient data is still fragmented in the majority of institutions: data of one patient may be distributed over paper records of various specialists and a variety of departmental and dedicated systems. Hence, the main challenges of today's developments are integration of data and support of SDE at the point of care.

At our department, we developed the computer-based patient record system ORCA (Open Record for CAre). In this paper, we will highlight its essential strategies from the developer and user point of view, i.e. the architecture and the user benefits, respectively. Since the barriers to integration involve infrastructure and are mainly of an organizational rather than technical nature, we will focus on SDE.

2. Architecture of ORCA

The goal of ORCA is not SDE itself, but the collection of data that are suitable for research, decision support, and shared care. It is not sufficient to make this goal possible. It must be practical. Besides unambiguous and consistent representation of data, ORCA must also accommodate different demands for contents and workflow to make it attractive and efficient for a variety of
specialists. In summary, the main requirements underlying the design of ORCA are:
1. Unambiguous representation of data, flexibility in scope and content, adaptability to different workflows, high level of maintainability.

Regarding the architecture of ORCA, we distinguish the data model and the structure of the software. The data model is essential for the first two goals. The structure of the software is important for accommodating different workflows and for the maintainability of ORCA.

2.1 The Data Model

Data in patient records can be divided into two main categories. The first category comprises data, for which the descriptors do not vary per specialty. Typical examples are laboratory test results and drug prescriptions; a test result is described by the type of test, the test value, and the units, whereas a drug prescription is characterized by the name of the drug, the dosage, and the frequency of intake. These domain-independent data can be represented in a straightforward, so-called ‘direct’ model with a direct mapping between the attributes in the database and the presentation in the interface [13]. Such a model is rigid in the sense that adaptations in content require changes at the query, functional, and interface level. This relative rigidity applies, in fact, to all models where structure and content are directly related. This is the case in many relational database applications. When there is, however, little need for adaptations, as is the case for domain-independent data, such a direct model is transparent and efficient.

The second category involves data that vary highly among different specialists. The most typical examples are patient history and physical examination. Each specialist focuses on different aspects. A direct model approach would require dedicated tables for different specialties, which will result in redundancy for overlapping data and a maintenance burden that would exponentially grow when scaling up. Furthermore, a direct approach cannot easily accommodate the recording of incidental findings. Hence, the recording of domain-dependent data requires a high degree of flexibility regarding both content and detail [14]. To accommodate flexible data entry ORCA uses a knowledge-driven approach. A knowledge base (KB) contains a controlled vocabulary of medical description knowledge, that defines which terms can be used and how they can be combined into medically meaningful descriptions [15]. The KB consists of a thesaurus of concepts and a directed graph, connecting these concepts. The children of a concept in the graph represent its descriptors. The KB is a graph because some concepts may apply in more than one context: dyspnea may be important in the context of both the cardiovascular and the respiratory system. The patient data, actually entered using the KB, is called the patient database. Its contents are trees, representing the order in which the user selected concepts from the KB. The advantage of both a graph and a tree is that the structure is independent of the contents: a change in content does not require changes to the data model or the interface.

2.1.1 The Knowledge Graph

The graph is a special application of conceptual graphs, in the sense that all semantics needed to represent the patient data are represented by the ordering of the concepts in the graph and not by the type of relationships between them. The relationships as found in traditional conceptual graphs are mapped in the ORCA KB to a concept and one of six types of relationships. For example, blood pressure – <has systolic pressure> – mmHg, is modeled in the ORCA KB as blood pressure – <has feature> – systolic blood pressure – <has unit> – mmHg. The relationships between the concepts only contain information to drive the data entry interface. Keeping in mind that the graph is solely used for support of data entry, this approach eliminates the need for additional knowledge to define how relationships, such as <has systolic pressure> should affect the behavior of the data entry interface. As the user traverses the knowledge graph, he will not be confronted with relationships: he will only see the concept sequences as they will be represented in the resulting patient data. The knowledge example above will result in: blood pressure – systolic pressure – mmHg.

The six relationships have different consequences for the user interface. The <has specialization> relationship is equivalent to a class-subclass relationship. The relationship <has feature> is used for qualitative descriptions of the parent. When the parent can be described quantitatively, the <has unit> children specify the applicable units of measure. Besides their own descriptors, <has specialization> children inherit the <has feature> and <has unit> children from their parent. A special relationship is <refers to>. This is used when a single concept may have more than one context. For example, micturition may be important in the context of the urogenital, cardiovascular, or endocrine system. It is not desirable, however, to have micturition descriptions appear at various places in the record. To preserve coherence of description, such concepts have a main parent via a <has feature> relationship and one or more <refers to> parents. The advantage is that the physician can access the concept via any appropriate path and will see immediately if and which findings are already entered.

The children of a concept in the knowledge graph constitute its descriptors in general, which means that the descriptors of a concept are context-independent. For example, a concept such as skin lesion could have the descriptors size, color, elevation, and contents, even though contents would only apply to blisters and pustules, and not to many other skin lesions.

The remaining two relationships come into effect during data entry when the user traverses the knowledge graph with the interactive browser. It is desirable that the user is not confronted with inappropriate descriptive options. The <preset choice> and <exclude choice> relationships tailor the generic descriptor sets to the context at hand. When two concepts are connected with the <preset choice> relationship, selection of the parent concept makes the child self-evident and selection is automatically performed by the system. For example, when the user has selected findings stomach – ulcer – location –, he will see only locations within the stomach,
whereas he sees only duodenum locations when he reached ulcer via findings duodenum. Similarly, `<exclude choice>` children are excluded from the descriptive options when the parent is selected.

Besides relationships between concept-pairs, concepts may have properties that pertain to the concept itself. Properties involve absence or presence, single or multiple instantiation within the same context, or mutual exclusivity of the descriptors. For example, ulcer has the ‘m’ (multiple) and ‘a’ (absent) properties because there may be multiple ulcers in the stomach or none. The ‘n’ (normal) property allows the user to specify a description of what he considers normal for that concept. Figure 1a shows a part of the knowledge graph with relationships and properties. Figure 1b shows which descriptors the user will see in the data entry interface.

2.1.2 Patient Data

Since all semantics are in the concepts and their order, the actual patient data are represented as a tree of instances of KB concepts only. A concept represents knowledge, i.e. an entity that can potentially be selected. An instance represents patient data: a concept that actually has been selected in the context of patient X. For example, what is modeled as ulcer – `<has feature>` location – `<has specialization>` stomach will produce the instance sequence ulcer – location – stomach, which conforms more to the traditional conceptual graph structure. The concepts in the graph and the instances in the data trees are represented by unique identifiers. Our knowledge-based approach is indirect in the sense that the identifiers of the instances need to be linked to their corresponding concepts in the KB thesaurus to become meaningful. The data representation in a tree format is the key to flexibility because it allows the separation of data structure and content. In the context of ORCA, a tree structure can represent a large variety of findings, irrespective of their type and detail.

2.1.3 Knowledge Modeling

The KB can be expanded and adjusted via an interactive knowledge editor.

Already recorded data, however, remain completely accessible: only the patient data tree and the thesaurus are needed to interpret the data. The ORCA software can be compared to a VCR and its data model to the VHS format. Just as the VCR can record and play different types of movies on VHS tape, the scope of ORCA can be expanded without changes to the data model and software.

2.2 The Software Architecture

The ORCA software consists of core components and application components. The core components encompass central software that communicates with the application components and the database. Examples of application components are patient administration, patient selection, clinical narrative, and medication. Three important strengths of the software architecture are the three-layer model of application components, the State Transition Manager (STM) in the core components, and the access to data from other systems.

2.2.1 The Three-layer Model

The three-layer approach reduces maintenance effort, because it allows for the reuse of code. The three layers are the query layer, the functional layer, and the interface layer. The query layer performs communication with the database and determines the data set with which the application component works. The functional layer performs operations on this data set, such as conversion of internal date and time formats, or the computation of derived data, such as age. The interface layer involves the presentation of the data to the user. Maintenance effort is reduced because different presentation of the same data only involves development of a new interface layer. The other two layers can be reused. Similarly, the
query layer can be re-used as long as changes to the functional layer involve the same data set. Some application components can be reused entirely: patient selection is used for the selection of a patient’s record as well as a patient’s administrative data.

2.2.2 The STM

The software architecture is modular in the sense that the application components have no awareness of each other. They are independent units of software. The STM controls the order in which the application components are activated via an application definition file with state transitions. An active application component is associated with a state, whereas a user-command that terminates the application component is associated with a transition. Upon selection of Close, OK, or a function button, the STM can read from the corresponding state transition in the definition file which application component needs to be activated next. As a result, ORCA can be adapted to different workflows by creating a file with different state transitions. In Fig. 2 the user selects ‘Patient Administration’ from the ‘Main’ window and is guided via ‘Patient Selection’ to ‘Patient Administration’.

2.2.3 Integration with Other Systems

For integration with existing systems, three components are distinguished: the plug, the drivers, and the savers. Drivers are specific for each external system. A driver addresses the external system in its own language and will convert the retrieved data to the format required in ORCA. The saver is specific per data category, such as laboratory test results or ECGs in the SCP standard. The saver stores all or only the requested part of the data in the data repository of ORCA. The plug is a generic software component, independent of the external system and the type of data. Based on knowledge in the ORCA database about the driver needed for each external system and the data category involved, the plug controls the issuing of commands to the external system and activates the appropriate saver for the returned data. The plug also monitors retries when requests for data were unsuccessful. Viewers for the respective data categories present the data from the external systems to the user as part of the patient record. Currently, savers and viewers are available for laboratory test results, ECG signals and data in SCP format, and angiographic images in Dicom format.

3. User Benefits of ORCA

The ORCA architecture offers a variety of benefits to the user. All benefits to be mentioned here are currently supported by the data model. Where the actual functional implementation has not yet been completed, we will indicate so.

3.1 Flexible SDE

ORCA supports SDE with a dynamic interface, instead of a rigid approach with a large set of fixed forms. ‘Dynamic’ refers to the fact that the descriptive options are context depen-
dent. The user traverses the knowledge network via a browser. For each selection made, new options appear in the appropriate medical context until the leaves of the network have been reached. Depending on which domains are covered in the KB, a specialist can enter findings in his own field as well as incidental findings in a different one. Descriptions may vary in detail, depending on the expertise of the user.

SDE in ORCA is fully integrated with free text to support data collection in both academic and regional settings. The user can associate free text with every selected concept. When SDE is not appropriate the option can be used at high level concepts, such as patient history. Free-text annotations can also be made at a specific level when the KB lacks expressiveness. Data entry can range from a predominantly free-text record to a predominantly structured one, which facilitates gradual migration from free-text entry to SDE. An example of ORCA’s clinical narrative window is shown in Fig. 3.

To enhance the efficiency of SDE and to tailor data entry to specific purposes, physicians can define views on the KB via the ‘form editor’. Using the browser, the physician can combine on a form those concepts that he routinely covers in a specific medical setting. Such forms can serve as a reminder and accelerate data entry since each concept on the form offers a shortcut to the browser. Many physicians have a routine way to screen certain parts of the body. As a result, records often contain statements, such as ‘heart normal’ or ‘abdomen normal’. The meaning of such expressions may differ per physician. ORCA allows physicians to define the meaning of their personal normal statements for later use. Such normal definitions provide ease of entry while preserving explicitness.

3.2 Patient Profile

The Patient profile offers an overview of the patient’s current status and direct access to the contents of the medical record. The profile offers a list of all past medical events regarding the patient. The list currently covers visits, laboratory test results, ECGs, and angiographic images. Furthermore, the patient profile shows the current medication, facts from the previous history, and risk factors. The information in the profile is intended to quickly refresh the physician’s memory of the patient. To view the patient’s status in the past, the user only needs to change the profile date and refresh the contents of the window. By selecting an item on the patient profile, the physician can view the corresponding data in full detail.

3.3 Multi-linguality

ORCA currently supports presentation of its contents in five languages: English, Dutch, German, French, and Italian. Greek is under development. This multi-lingual support includes both the structured data and the labels of the interface itself. The language can be switched during run time. Free text remains in the original language. Since all structured data are based on the same KB, physicians can read non-native data in their own language.

3.4 Shared Records

The ORCA data model allows for sharing of records by different physicians treating the same patient. All patient data are linked with their author, specialty, and department. Via constraints on these links, the user can have an author view, specialty view, or department view. Retrieval without constraints on these links produces the overall view. Currently, only the default view, i.e., the specialty view is being used. In overall view, the patient profile lists all data regarding the patient, including the complete previous history and all medications. This view may reduce undesirable combinations of drugs and superfluous requests for additional examinations.

4. Discussion

Integration is an effective strategy to enhance the benefits of a CPR, because data come from other systems and need not be entered by clinicians [16]. Integration mainly requires the proper infrastructure, standardization, and technical effort. Physicians are usually not the limiting factor. Clinicians’ narratives, however, play an important part in virtually all medical specialties. Without them, a multi-media CPR can never be more than a supplement to the paper record. For a long time, data entry by physicians was not considered feasible and many systems still rely on data forms, that are transcribed and entered later on by clerical personnel. Yet, typing is not necessarily as great a barrier for clinicians as many believe. At the Beth Israel Hospital in Boston, clinicians voluntarily began to type free-text progress notes in The Outpatient Medical Record (OMR) [17]. Dual charting, however (i.e., the use of both a CPR and the paper record) bears the risk of overlooking data. Physicians tend to rely on the source that is found to be most complete, but one can never guarantee instantaneous completeness of either one. Therefore, the CPR must fully replace the paper record to reach its full potential [1, 17].

Although the entry of progress notes is a step towards a more complete CPR, free text is not suitable for automated interpretation and data analysis [18]. Although natural language processing, i.e., extracting structured data from free text, has little impact on a physician’s working style, several studies have demonstrated that SDE has a far greater potential for the quality and completeness of the data being collected [8, 19, 20]. When clinicians are willing to type, is SDE a feasible goal? There is continuing effort at making SDE feasible within the time-pressured environment of the practicing physician [9–12, 21, 22]. Some of the resulting applications are dedicated for a specific domain and have a limited scope [11, 21]. Pen & Pad mainly focuses on use by general practitioners. IMR-E also has a large scope, but is primarily based on complaint descriptions. What these SDE applications have in common is that they offer predefined options for description, based on a controlled vocabulary. They are dynamic in the sense that the descriptors vary per context. Some are completely dynamic in the sense that every selection of a descriptive option determines what the next options are. Others use a more template-based approach where a set of descriptors and modifiers are offered on the basis of a...
selected complaint or finding. Such templates look like complaint-specific forms. These forms are intended to find the optimum between flexibility and efficiency of data entry. Depending on the model underlying the vocabulary, the depth or detail of description may be fixed or variable: some models are based on fixed-depth attribute-modifier models [23], whereas others use unlimited conceptual graphs [11, 22]. For some systems it is not clear to what extent the underlying design is generic and could be applied to domains other than those known from the literature. The ORCA model is designed for application in a variety of medical domains and without limitations on detail of description. The interface promotes SDE, but this is not mandatory. Free-text entry is fully integrated to permit gradual migration from a predominantly free-text record to a more structured record.

Preliminary evaluation of ORCA involved an experiment in which clinicians were asked to enter data and answer questions about cases in the CPR [24]. As strengths of ORCA they mentioned the suitability of the structured data for research and shared care, and the fact that retrieval of data was faster than from paper records. Yet, SDE was still found to be complex and time-consuming. Part of this experience can be attributed to the fact that none of the participating physicians had much experience with the system. In addition, the experimental setting precluded the advantage of ORCA’s features to enhance data entry, such as shortcut forms and normal definitions. SDE does not necessarily require extra time [19]. In the Catharina Hospital in Eindhoven (the Netherlands), the ICIS system in the ICU unit is found to be more efficient than recording on paper records. However, the efficiency of data entry did not yet meet the requirements of routine care. The main challenge is to better anticipate the data set to be entered. This is straightforward in protocol-based care, but in less structured settings such as the outpatient clinic, this requires multiple views from a problem-based perspective. Current research focuses on the ways to represent problem-oriented views for data entry.

Knowledge modeling has been done in the domains of radiology, neurology, and pathology, proving the versatility of the knowledge-based approach. Projects with pediatrics and andrology have recently started.

5. Current Status and Future Developments

At present, ORCA has been installed at six European centers in the context of the 1IC4 (Integration and Communication for the Continuity of Cardiac Care) project. Preliminary evaluation results show that clinicians see the advantages of SDE for research, shared records, and decision support. Retrieval of data was experienced as faster than retrieval of paper-based data. However, the efficiency of data entry did not yet meet the requirements of routine care. The main challenge is to better anticipate the data set to be entered. This is straightforward in protocol-based care, but in less structured settings such as the outpatient clinic, this requires multiple views from a problem-based perspective. Current research focuses on the ways to represent problem-oriented views for data entry.

Knowledge modeling has been done in the domains of radiology, neurology, and pathology, proving the versatility of the knowledge-based approach. Projects with pediatrics and andrology have recently started.

Acknowledgments

This research has partly been funded by the Fourth Framework Programme on Health Tele-matics C7/ of DGXIII, European Community, Brussels.

REFERENCES

8. Whiting-O’Keefe QE, Simborg DW, Epstein WV, Warger A. A computerized summary medical record system can provide more information than the standard medical record. JAMA 1985; 254: 1185-92.


24. Deliverable 4 of I4C (Integration and Communication for the Continuity of Cardiac Care) of the fourth Framework program of the European Community, 1999.


Address of the authors:
A. M. van Ginneken MD, PhD, Dept. of Medical Informatics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
E-mail: vanginneken@mi.fgg.eur.nl