

Hadron therapy information sharing prototype

Faustin Laurentiu ROMAN^{1,2,3,*}, Daniel ABLER^{1,4}, Vassiliki KANELLOPOULOS^{1,5},
Gabriel AMOROS², Jim DAVIES^{4,6}, Manjit DOSANJH¹, Raj JENA⁷, Norman KIRKBY⁵,
Ken PEACH^{4,6} and Jose SALT²

¹CERN, European Organization for Nuclear Research, CH-1211 Geneva 23, Switzerland

²IFIC, Instituto de Física Corpuscular, Apartado de Correos 22085, E-46071 Valencia, Spain

³Present address: EBG MedAustron GmbH, Marie Curie-Strasse 5, A-2700 Wiener Neustadt, Austria

⁴Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

⁵Faculty of Engineering & Physical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK

⁶PTCRI, Particle Therapy Cancer Research Institute, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

⁷Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, CB2 0QQ, UK

*Corresponding author: Faustin Laurentiu Roman, EBG MedAustron GmbH, Marie Curie-Strasse 5, A-2700 Wiener Neustadt, Austria. Tel: +43-2622-26-100-621; Fax: +43-2622-26-100-119; Email: faustin.roman@medaustron.at

(Received 12 January 2013; revised 15 March 2013; accepted 21 March 2013)

The European PARTNER project developed a prototypical system for sharing hadron therapy data. This system allows doctors and patients to record and report treatment-related events during and after hadron therapy. It presents doctors and statisticians with an integrated view of adverse events across institutions, using open-source components for data federation, semantics, and analysis. There is a particular emphasis upon semantic consistency, achieved through intelligent, annotated form designs. The system as presented is ready for use in a clinical setting, and amenable to further customization. The essential contribution of the work reported here lies in the novel data integration and reporting methods, as well as the approach to software sustainability achieved through the use of community-supported open-source components.

Keywords: hadron therapy; proton therapy; data federation; web portal; eHealth; cancer informatics

INTRODUCTION

There is a need for increased data sharing in cancer care and cancer research [1, 2]. In the case of hadron therapy, this need is particularly pronounced. Also known as particle or ion beam therapy, it involves the innovative use of protons and carbon ions [3], and while this promises significant advantages, more information is required to support adoption, to determine suitability, and to support treatment planning. Furthermore, there are relatively few treatment centres, and patients often cross national boundaries for treatment. There is a requirement for effective follow-up to compare outcomes and establish efficacy [4], and a European initiative has been established to provide this [5, 6].

An analysis of the hadron therapy domain, considering the requirement for data interoperability and the achievements of existing eHealth initiatives, has informed the design of a ‘Hadron therapy Information Sharing Prototype’ (HISP), as a

gateway to patient information held in multiple hospital databases and a means of supporting patient follow-up in multi-centre clinical studies [7, 8].

To demonstrate the functionality of the system, we focus on an adverse-event-reporting scenario—a key component of comprehensive patient follow-up. The reporting of adverse, treatment-related events is part of the patient management process, beginning with the initial visit where a baseline is assessed, continuing during treatment with each clinical review, and then at intervals during follow-up. This demonstration covers the main aspects of system functionality: patients and doctors report an adverse event as structured, coded information; this information is integrated into medical records at treatment centres, and is then available, across a distributed architecture, under role-based access control.

The prototype was developed within the PARTNER Project, as a collaborative effort between three Marie-Curie

Early Stage Researchers and four institutions: CERN, IFIC, the University of Oxford, and the University of Surrey [5]. This paper reports upon the design of the architecture and principal components, as well as upon the benefits and limitations of the existing implementation.

MATERIALS AND METHODS

The system provides a single access point for patient information distributed across multiple hospital databases. At the design level, it can be seen as a generic server infrastructure, with services for data integration and presentation, and other services for reporting and analysis.

System architecture

The particular challenges within this domain include (i) a lack of interoperability between existing clinical care systems; (ii) the distribution of information pertaining to an individual patient across multiple systems; and (iii) a lack of consistent and amenable policies for data access. These challenges had a profound influence upon the system design, the data model, and the development process.

The architectural design was based upon the storyline or pathway of an individual patient undergoing hadron therapy [9]. In this storyline, data is collected in two countries, from different actors: doctors and patient. Data is held within local networks, but made available to external services. The prototypical infrastructure was realized on four servers distributed across two sites (Fig. 1) [10, 11]:

- CERN server 1: presentation services, user interface
- CERN server 2: integration services
- CERN server 3: metadata and reporting services, data capture
- IFIC server 4: database located at second site.

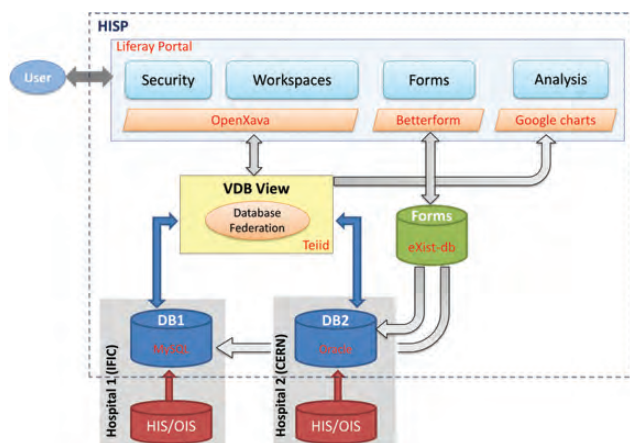


Fig. 1. HISP architecture: data stored in DB1 and DB2 is held locally but available remotely.

Databases

The hospital data repositories, Hospital or Oncology Information Systems (HIS/OIS), export data to local HISP databases remaining under local governance (DB1 and DB2). The basic data model adopted for this scenario represents a ‘care summary’ view of the data, suitable for use across institutions. Raw images, laboratory reports, and treatment plans are not shared in this scenario. Sample data to validate the system was generated from published clinical trials, and imported into the hospital databases for testing purposes.

The model [12] comprises four tables of data: patient information (e.g. demographics, history); tumour information (e.g. type, staging); treatment information (e.g. dose and beam quality); and adverse-events information (e.g. scoring system, adverse event, grade, time of onset). The data in the first three tables would be imported automatically from hospital systems; only the adverse-events data needs to be acquired separately. The structure of the model allows for the submission of patient- and doctor-reported event information to different data standards. In the example schema, however, any reported event has to be represented as a combination of ‘adverse-event name’, ‘severity score’ and ‘scoring system’. For the inclusion of patient-reported outcome measures, this would require an initial transformation step in the reporting services, anticipated in the architecture design.

One part of the generated patient data was stored in the IFIC ‘MySQL’ server (DB1); the other was stored in a CERN ‘Oracle’ database instance (DB2) [13, 14]. As ‘Oracle’ databases, often used in hospital environments, have different characteristics to ‘MySQL’ databases (operations, schema and access), this allowed us to test low-level interoperability and integration functions.

Data integration

The architecture relies upon ‘data federation’ to provide an integrated view across heterogeneous data sources, presenting their aggregation as a virtual database. This virtual database does not contain the data itself, but instead holds information about the data locations, types, and access procedures. The data remains within the source organisations, which are able to determine the extent to which it may be accessed across the system.

‘JBoss Teiid’ was used as the federation platform [15], hosted on CERN server 2, and used to integrate the two databases, ‘MySQL’ and ‘Oracle’, to produce a virtual database (VDB) view. The view was restricted to address the particular requirements of the adverse-event-reporting scenario [10].

Portal

The user interface for the system was based upon the ‘Liferay portal’, an enterprise-level, open-source standards-compliant web platform [16]. The medical domain has

particular requirements for standards, certification, and support; this technology addressed these, while providing a range of useful, packaged functionality. For the interface, we used ‘Openxava’: a ‘Java’-based framework for the rapid development of internet applications [17], compatible with ‘Liferay’. The combined framework allowed us to produce interfaces that were both easy to use and easy to develop.

Security

Three primary roles are defined for authorization (doctor, patient, researcher), together with two further, secondary roles (new user and administrator) [10]. ‘Doctor’ is the most privileged role, affording access to all patient data within the federation. ‘Researcher’ allows a simple statistical analysis. ‘Patient’ affords access to the follow-up forms. A given user can be assigned multiple roles, depending upon the context of current activity.

Communication with the HISP portal (server) is secured using the Transport Layer Security (TLS) version 1.2 encryption protocol. Users can login via password-based authentication or public key certificates loaded into the client browser: the second approach, suitable for use within a controlled hospital environment, facilitates access and provides increased security overall [18].

Data reporting

The data reporting functionality was designed for web-based structured reporting of medical findings, with minimum standards of documentation, facilitating the re-use of data across contexts.

Metadata services

To facilitate re-use, the intended meaning of the data must be documented and be available as ‘metadata’ in a computable form. Metadata registries serve as common reference points for metadata standards, and provide services for metadata registration, curation and versioning. ISO/IEC 11179 is a standard for metadata registries, addressing ‘the semantics of data, the representation of data, and the registration of the descriptions of that data’ to promote common understanding, harmonization, and re-use [19]. A key part of this standard is the notion of a common data element, a fundamental unit of metadata that may be applied to data collected or managed in a different context. A data element is intended to provide a full description of the meaning of a given observation. To provide the required metadata support for the adverse-event scenario, we used the ‘CancerGrid’ metadata registry (cgMDR) [20], an open-source implementation of the ISO/IEC 11179 standard.

Data reporting services

The reporting services had to be widely accessible: to provide verification of data entry prior to submission, and to allow for changes in requirements to be easily

accommodated through the versioning of data-capture instruments and data schemas. This latter aspect is particularly important for early-phase clinical studies, where the set of observations to be made is not finalized at the outset of the study. It is important also in the context of new or evolving reporting instruments and standards, e.g. where new measures are being devised for the reporting of ‘subjective’ adverse events after therapy.

The ‘Extensible Markup Language’ (XML) [21] was chosen as the serialization format for submitted data, reflecting a flexible document-centric approach in which every form submission corresponds to a single observation record. The validity of any submitted record can be verified using an XML schema. By annotating the schema files with SAWSDL references [22] pointing to the respective data elements registered in the metadata registry, collected data is prepared for future data-integration scenarios.

Data-reporting forms were developed in ‘XForms’ [23], an XML-based standard recommended by the World Wide Web Consortium (W3C) for the declarative definition of web forms. We used ‘betterFORMS’, a server-side implementation of XForms [24], to make the forms accessible using standard web technologies. Validation rules within the form specification are used to check data prior to storage in a native XML database: ‘eXist-db’ [25], was chosen for this purpose.

The identity of the submitter (patient or doctor and respective credentials), the subject (patient and credentials), and the associated treatment centre is verified at form submission. A user-defined extract of the reported data, possibly involving data transformations to comply with the target schema, is then transmitted into the local database (DB1 or DB2 in Fig. 1) and becomes available to federated queries. In the prototype implementation, this functionality is provided by a ‘ModPython’ [26] script, retrieving the submitted XML record, performing the necessary data transformations, and generating the query statements to insert the data into the database.

All of the above services are hosted on CERN server 3. Data-reporting forms were created for patient-reported outcome measures, based on existing questionnaires [27]. Forms were created for the submission of adverse-event reports, based on reporting standards such as CTCAE, RTOG and SomaLent [28–31].

Analysis service

The federated data can be used to provide a range of analysis services. Depending upon the user’s role, these can be customized to address specific needs. We used PHP and the Google charts API [32] to build a simple proof-of-concept report. This queries the VDB anonymously to produce statistical information about the cumulated adverse events: for example, a piechart display of the total numbers of the five most common adverse events. A range

of parameters for the queries can be selected using drop-down menus.

RESULTS

The HISP PARTNER prototype system demonstrates a particular approach to the acquisition and re-use of relevant medical data. The following features are seen as benefits or advantages of the approach:

- the ability to query data across a federation of heterogeneous database systems
- the availability of an easy-to-use portal interface with role-based access control
- the use of semantic web and metadata technologies to facilitate semantic annotation of data
- the support for versioning of patient-reported outcome and objective adverse-event measures.

DISCUSSION

Although existing solutions for clinical data management—such as OpenClinica [33], caAERS [34], or the particle database system developed within the European ULICE-framework [35]—provide functionalities that the HISP prototype does not support, the system serves well to demonstrate a range of different and important features.

Rather than employing a traditional ‘data warehousing’ approach [35], HISP relies on ‘data federation’ to provide an integrated view across heterogeneous data sources. A warehousing approach requires a fixed model, less amenable to change, and could lead to concerns over data control and privacy. This was a particularly important consideration given the evolving nature of the science and medical practice, as well as the concrete specification of governance constraints upon medical data (typically, ‘the data stays within the hospital; only suitably abstracted versions of the data may be made available externally, and then only for clearly specified purposes’).

Where reporting scenarios involve multiple institutions, and where data collection may take place over a period of years, clear documentation of reporting intent is needed for subsequent, correct interpretation and re-use. As the scale of data collection increases, the importance of machine-readable documentation—computable metadata—increases also. The advantages of semantic annotation based upon metadata standards have also been demonstrated in [36]. An important feature of the HISP design is that annotation is part of the form design process, ensuring that data is associated, automatically, with a computable representation of its meaning at the point of collection.

A model-driven approach is used to automate the process of form implementation, minimizing the development effort

involved [37], improving quality, and reducing the cost of validation [38]. Other reporting systems frequently provide the means of facilitating the form design process by providing ‘form builders’: some of these are simplistic [39]; others treat each form as an indivisible whole [33, 40], preventing comparison and re-use at a data-element or question level. As different conditions, different questions, and even different individuals may be presented with different forms [41], the ability to compare designs at a question level is essential for data re-use and integration. The ability to re-use designs at this level is equally essential for the harmonization of reporting standards and procedures. Lessons learned from the prototype development have informed the design of comprehensive metamodel, or domain-specific language, for form specification [42].

The use of open-source components is incidental to the design of the system, but makes an important contribution to its extensibility and potential for adoption. The purpose of the prototype is to establish technical feasibility, and significant extension would be required for wider deployment outside the context of the originating project: for example, a collection of standard interfaces to hospital databases and messaging systems. However, the system as it stands serves as a complete demonstration for the proposed approach, as well as a platform for further development.

ACKNOWLEDGEMENTS

The authors are grateful to Prof. B. Jones (Gray Inst., PTCRi) and Prof. R. Mayer (EBG MedAustron GmbH), as well as the hospitals of Oxford and Valencia for discussions regarding all medical aspects of the project. This project would not have been possible without the assistance of M. Cirilli, A. Di Girolamo, H.F. Hoffmann and A. Valassi (CERN). Further, the authors would like to thank the CERN IT-ES group, as well as S. Harris and C. Crichton (University of Oxford, PTCRi) for technical support and numerous valuable discussions.

FUNDING

This work was supported by a Marie Curie Initial Training Fellowship of the European Community’s Seventh Framework Programme under contract number [PITN-GA-2008-215840-PARTNER].

REFERENCES

1. Sommer J. The delay in sharing research data is costing lives. *Nat Med* 2010;**16**:744.
2. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;**88**:462–66.
3. Wilson RR. Radiological use of fast protons. *Radiology* 1946;**47**:487–91.

4. Lodge M, Pijls-Johannesma M, Stirk L *et al.* A systematic literature review of the clinical and cost-effectiveness of hadron therapy in cancer. *Radiother Oncol* 2007;**83**:110–22.
5. Particle Training Network for European Radiotherapy (PARTNER) Project, FP7 Grant agreement no. 215840-2. <http://cern.ch/partner> (10 December 2012, date last accessed).
6. Union of Light Ion Centres in Europe (ULICE) Project, FP7 Grant agreement no. 228436. <http://cern.ch/ulice> (10 December 2012, date last accessed).
7. Roman FL. Health grids: overview, added-value. PARTNER Work Package 22—Deliverable 1. Technical report. *CERN*, 2009.
8. Abler D. Data integration for the PARTNER hadron therapy information sharing platform (HISP). PARTNER Work Package 23—Deliverable 1. Technical report. *CERN*, 2010.
9. Roman FL, Abler D. Services of the Hadron therapy Information Sharing Platform. Legal and Ethical implications. PARTNER Work Package 22 and 23 – Deliverable 2. Technical report. *CERN*, 2010.
10. Roman FL. Data-driven platform for hadron therapy. PARTNER Work Package 22 – Deliverable 3. Technical report. *CERN*, 2011.
11. Abler D. Data capturing services for charged particle therapy. PARTNER Work Package 23 – Deliverable 3. Technical report. *CERN*, 2011.
12. Kanellopoulos V. A Data Model for the Follow-Up and Evaluation of Side Effects. PARTNER Work Package 24 – Deliverable 4. Technical report. *CERN*, 2010.
13. MySQL database. <http://www.mysql.com/> (10 December 2012, date last accessed).
14. Oracle database. <http://www.oracle.com/> (10 December 2012, date last accessed).
15. JBoss Teiid. <http://www.jboss.org/teiid/> (10 December 2012, date last accessed).
16. Liferay Portal. <http://www.liferay.com> (10 December 2012, date last accessed).
17. Openxava. <http://openxava.org/web/guest> (10 December 2012, date last accessed).
18. Java SE Security. <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-136007.html> (10 December 2012, date last accessed).
19. ISO/IEC 11179-1: 2004(E). <http://metadata-standards.org/11179/> (10 December 2012, date last accessed).
20. CancerGrid metadata registry (cgMDR). http://www.cancergrid.org/index.php?option=com_content&id=8:mdrarticle (10 December 2012, date last accessed).
21. Extensible Markup Language(XML). <http://www.w3.org/XML/> (10 December 2012, date last accessed).
22. W3C Semantic Annotations for WSDL and XML Schema. <http://www.w3.org/TR/sawSDL/> (10 December 2012, date last accessed).
23. W3C XForms 1.1. <http://www.w3.org/TR/xforms11/> (10 December 2012, date last accessed).
24. betterFORM. <http://www.betterform.de/en/index.html> (10 December 2012, date last accessed).
25. eXist-db database. <http://exist-db.org/exist/index.xml> (10 December 2012, date last accessed).
26. Mod_Python. <http://www.modpython.org/> (10 December 2012, date last accessed).
27. Patient reported outcome measures: head and neck radiotherapy patient questionnaire. www.christie.nhs.uk/media/106098/Raphnsubj.pdf (10 December 2012, date last accessed).
28. Acute Radiation Morbidity Scoring Criteria. <http://www.rtog.org/ResearchAssociates/AdverseEventReporting/AcuteRadiationMorbidityScoringCriteria.aspx> (10 December 2012, date last accessed).
29. RTOG/EORTC Late Radiation Morbidity Scoring Schema. <http://www.rtog.org/ResearchAssociates/AdverseEventReporting/RTOGEORTCLateRadiationMorbidityScoringSchema.aspx> (10 December 2012, date last accessed).
30. SomaLent. http://www.rtog.org/LinkClick.aspx?fileticket=_rSiZbdkeXs%3d&tabid=140 (10 December 2012, date last accessed).
31. Common Terminology Criteria for Adverse Events (CTCAE). http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14.owl (10 December 2012, date last accessed).
32. Google charts API. <https://developers.google.com/chart/> (10 December 2012, date last accessed).
33. OpenClinica. <https://community.openclinica.com/> (10 December 2012, date last accessed).
34. Adverse event reporting system (caAERS). <http://cabig.cancer.gov/solutions/applications/caaers/> (10 December 2012, date last accessed).
35. Kessel KA, Bougaf N, Bohn C *et al.* Connection of European particle therapy centers and generation of a common particle database system within the European ULICE-framework. *Radiat Oncol* 2012;**7**:115.
36. Papatheodorou I, Crichton C, Morris L *et al.* A metadata approach for clinical data management in translational genomics studies in breast cancer. *BMC Med Genomics* 2009;**2**:66.
37. Calinescu R, Harris S, Gibbons J *et al.* Model-driven architecture for cancer research. In: *Proceedings of the 5th IEEE International Conference on Software Engineering and Formal Methods, London*, 2007.
38. Davies J, Gibbons J, Harris S *et al.* Model-driven support for a vaccine study in Kathmandu. Microsoft eScience Workshop, 2009.
39. caBIG FormBuilder. <https://cabig-stage.nci.nih.gov/community/tools/FormBuilder> (10 December 2012, date last accessed).
40. REDCap. Research Electronic Data Capture. <http://project-redcap.org/> (10 December 2012, date last accessed).
41. Richesson RL, Nadkarni P. Data standards for clinical research data collection forms: current status and challenges. *JAMIA* 2011;**18**:341–46.
42. Abler D, Crichton C, Welch J *et al.* Models for Forms. Proceedings of the Compilation of the Co-located Workshops SPLASH 2011. *ACM* 2011;13–18.