Public Health and Informatics
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210169

FhirSpark – Implementing a Mediation Layer to Bring FHIR to the cBioPortal for Cancer Genomics

Niklas REIMER[a,1], Philipp UNBERATH[b], Hauke BUSCH[a,*] and Josef INGENERF[c,d,*]
[a] *Group for Medical Systems Biology, Lübeck Institute of Experimental Dermatology, University of Lübeck, Germany*
[b] *Department of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[c] *IT Center for Clinical Research, University of Lübeck, Germany*
[d] *Institute of Medical Informatics, University of Lübeck, Germany*

Abstract. cBioPortal is a commonly used data warehousing solution for genomic cancer studies. The software is being extended for patient care application in a molecular tumor board by the MIRACUM consortium within the Medical Informatics Initiative Germany. A key feature for this use case is the ability to enter therapy recommendations for individual patients, which requires interoperability with the hospital information system. A RESTful interface between cBioPortal and an external mediation layer was selected from the different implementation options. It achieves interoperability by using a FHIR capable server to store data and applying the HL7 FHIR Genomics Reporting implementation guide. For systems not supporting the FHIR standard, the well-established HL7 Version 2 standard is available as a fallback export format.

Keywords. cBioPortal, HL7 FHIR, Health Information Interoperability, Precision Medicine

1. Introduction

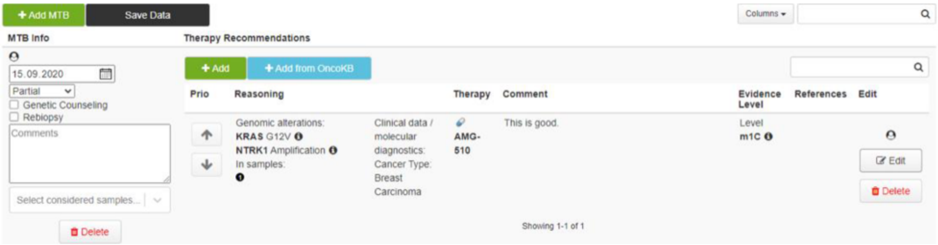
Precision medicine is an emerging field and the future of patient care in the healthcare sector, especially when using molecular data for personalized cancer treatment, such as whole genome or panel sequencing, transcriptomics, or immunohistochemistry. Personalized oncology approaches are currently being implemented within so-called molecular tumor boards (MTBs), which base their treatment recommendations on the individual molecular identity of each tumor, which is in contrast to traditional entity-specific tumor boards. Naturally, the implementation of such an MTB requires an interdisciplinary team of oncologists, pathologists, and bioinformaticians. Providing the necessary IT support is one of the missions of the MIRACUM consortium [1] within the Medical Informatics Initiative Germany (MI-I) [2].

[1] Corresponding Author, Niklas Reimer, Group for Medical Systems Biology, Lübeck Institute of Experimental Dermatology, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; E-mail: n.reimer@uni-luebeck.de.

[*] These authors contributed equally to this work.

One task is to adapt and integrate the data warehouse for clinical cancer studies, cBioPortal, for the routine patient care within an MTB. cBioPortal is an open-source tool mainly developed by the Memorial Sloan Kettering Cancer Center that provides integrated access and interfaces to more than 300 public cancer studies [3, 4]. As it is designed with a data warehousing approach in mind, it lacks various features for routine use in an MTB, as evaluated in a requirements analysis by Buechner et al. [5]. Adding such features is possible, but not straight forward as pointed out after a more in-depth analysis of the cBioPortal architecture [6]. The study revealed that cBioPortal is split into a frontend and a backend part but has no plugin concept. As a result, updates from the main project will often conflict with self-made modifications to the code base, so that continuous adjustments will be required.

An essential functionality identified by Buechner et al. [5] is documenting therapy recommendations for patients directly in cBioPortal. This feature is particularly challenging as its implementation impacts not only the codebase itself but also the persistence layer to store new recommendations. Additionally, the application should be easy to integrate with existing hospital information systems.



MTB Info

15.09.2020

Partial

□ Genetic Counseling

□ Rebiopsy

Comments

Select considered samples...

Therapy Recommendations

Prio	Reasoning		Therapy	Comment	Evidence Level	References	Edit
↑ ↓	Genomic alterations: KRAS G12V NTRK1 Amplification In samples:	Clinical data / molecular diagnostics: Cancer Type: Breast Carcinoma	AMG-510	This is good.	Level m1C		

Showing 1-1 of 1

Figure 1. Implementation of therapy recommendation form described by Buechner et al. [5]. The user can enter multiple therapy recommendations in multiple sessions of the tumor board. A single recommendation is a drug target based on selected genetic alterations, corresponding clinical parameters, comments, and an evidence level. The evidence level is usually backed by relevant publications.

2. Material and Methods

The therapy recommendation form had already been added as a front-end prototype to a forked cBioPortal version[2] as shown in Figure 1. This version has been deployed and extended to interact with a RESTful interface using the create, read, update, and delete (CRUD) pattern. As the user interface of cBioPortal is written in TypeScript, the model for therapy recommendations is backed by type definitions and can be easily exported to objects using the JavaScript Object Notation (JSON). However, this format is not conformant to any standard and is driven by the prior requirements analysis of Buechner et al. [5] and type definitions of the contained TypeScript objects.

2.1. Mapping to a standardized data model

In order to standardize the representation of the JSON export, a mapping to Fast Healthcare Interoperability Resources (FHIR) as part of the Health Level 7 (HL7) standard family is envisioned. The HL7 FHIR standard delivers implementation guides

[2] https://github.com/buschlab/MIRACUM-cbioportal

for use cases like *Genomics Reporting* which are used to represent therapy recommendations as a *Genomic Report* that comes along with therapeutic implications [7].

Usually, FHIR is deployed using a complete FHIR server with its own persistence layer, while legacy applications are connected using synchronization mechanisms [8]. Another approach is an implementation on top of a legacy data model with a translation layer between the FHIR interfaces and the database. Our use case is based on a legacy data model represented by a JSON schema but without a database implementation. This allowed us to combine both approaches and only store the FHIR resources, while the JSON representation is being generated on the fly when it is requested.

2.2. Integration with cBioPortal

Users of cBioPortal can be either anonymous or authorized using an identity provider, which also allows hosting private datasets where access is granted to specified users or groups.

This should also be applied to the envisioned extension for therapy recommendations. Otherwise, sensitive information could be accessed and modified by unauthorized users. Especially for implementations decoupled from the cBioPortal codebase for better compatibility with updates, this is quite challenging. Those will not be able to use the internal routines of cBioPortal to check whether a user is authorized or not.

2.3. Evaluation methods

To ensure that the generated FHIR resources are valid and fulfill all constraints, the FHIR reference implementation HAPI FHIR was set up as a server and the *Genomics Reporting* implementation guide was imported. The generated HL7 Version 2 messages were validated using the built-in validator of HAPI TestPanel.

Test cases are set up to ensure a correct bijective projection between JSON objects and FHIR resources is correct. They transform test datasets from JSON objects to FHIR resources, upload them to the FHIR server, query them and transform them back into the JSON format. The resulting output JSON is compared to the test input.

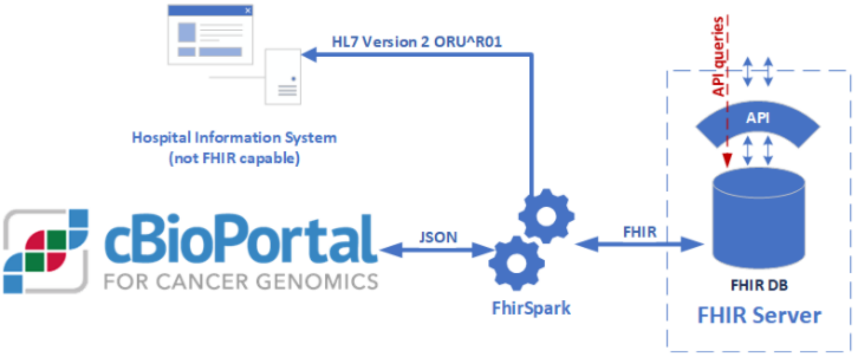


Figure 2. Proposal for the integration of cBioPortal in the clinical infrastructure. FhirSpark acts as a mediation layer between cBioPortal and a FHIR capable server. Systems that do not support FHIR can also be connected using HL7 Version 2 messages. Instead of having a legacy database without FHIR that is either synchronized or mapped to FHIR as described by Smirnov [8], FHIR is used as a native storage format.

3. Results

FhirSpark is implemented to provide a Webserver with RESTful services for access by cBioPortal. Incoming data is mapped to Java objects after a validation step. The corresponding Java classes were generated from the JSON schema. Internally, FhirSpark acts as a mediation layer between the RESTful interface and a FHIR-compliant server while being completely stateless. Saved therapy recommendations are being transformed into FHIR resources, compliant with the Genomics Reporting implementation guide.

As some values come in with a *display* value only, they will be enriched with the corresponding *code* from the originating databases in order to fulfill the coding constraints of the implementation guide. For example, the gene symbol *TP53* is a display value that must be committed with its corresponding code *HGNC:11998* as a *gene-studied* component according to the Structure Definition *Variant*. The same is true for publications that are being sent with a PubMed id only. As shown in Figure 2, our approach does not store the incoming data in a traditional relational database in favor of persistence using native FHIR resources. Nevertheless, this comes with the tradeoff that all attributes must be mapped to FHIR resources to be stored. This is achieved by using FHIR value sets in the therapy recommendation form, if applicable, or by additional enrichments as described for PubMed publications.

The resulting resources will be collected in a bundle that can be processed by the server at once. For that, conditional updates of the FHIR specification were used so that an existing resource can be reidentified and updated even though all resources inside the bundle are only linked through unique temporary identifiers. This avoids storing the data in multiple chunks.

For legacy hospital information systems without support for the FHIR standard, a fallback option is offered to send finalized reports in the HL7 Version 2 format to a predefined target using the *ORU^R01* trigger event for unsolicited observation messages. This is possible because the implementation guide relies on panels defined in the Logical Observation Identifiers Names and Codes (LOINC) terminology. The adaption of such panels with HL7 Version 2 has already been shown with the *Genetic Variation Model* implementation guide for HL7 Version 2.

As therapy recommendations in MTBs focus on targeting genetic alterations through context-specific drug treatment, all drug-target combinations, as well as associated publications and studies, constitute a valuable knowledgebase. For this reason, two additional endpoints were implemented to query therapy recommendations and publications based on an array of genetic alterations. For this, the existing type definitions were reused.

FHIR Search is used to query the data from the FHIR server for the transformation back to the JSON format, as well as the RESTful endpoints for therapy recommendations and publications described above. For better performance, the capabilities of the *include* and *rev-include* attributes were used, so that referenced resources will be included instead of being queried using multiple requests causing additional latency.

To address the authorization, FhirSpark is deployed together with a reverse proxy serving both cBioPortal and FhirSpark. This triggers the web browser to send the authentication token also to FhirSpark, which allows forwarding it to cBioPortal to check whether access should be granted or not.

The source code of the implementation is available on GitHub[3].

[3] https://github.com/buschlab/fhirspark

4. Conclusion and Outlook

In this work, we presented a novel approach that extends the widely used cBioPortal to store therapy recommendations from molecular tumor boards using the FHIR standard. Although this add-on function is tightly coupled to cBioPortal, the presented implementation is not directly integrated into its source code while avoiding redundant data storage. The proposed FhirSpark application acts as a mediation layer to offer a RESTful interface for cBioPortal. It is able to provide a bijective translation between FHIR resources and the JSON format for cBioPortal.

The cBioPortal application will be made available on the technical platforms of the MIRACUM and HiGHmed [9] consortia of the MI-I, although with FHIR and openEHR different they rely on different specifications to store clinical data [10].

Furthermore, a future extension will integrate additional follow-up data and investigate whether this structure can also be integrated into the extract, transform, load process when importing data into cBioPortal.

Acknowledgment

The project is funded by the German Federal Ministry of Education and Research (BMBF), Grand IDs 01ZZ1802Z (HiGHmed) and 01ZZ1801A (MIRACUM).

The authors acknowledge support through the HiGHmed and MIRACUM consortia as part of the Medical Informatics Initiative Germany.

References

[1] Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods of Information in Medicine. 2018jul;57(S 01):82–91.

[2] Semler S, Wissing F, Heyder R. German Medical Informatics Initiative. Methods of Information in Medicine. 2018 jul;57(S01):50–56.

[3] Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr;6(269):l1.

[4] Cerami E, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery. 2012;2(5):401–404.

[5] Buechner P, et al. Requirements Analysis and Specification for a Molecular Tumor Board Platform Based on cBioPortal. Diagnostics. 2020; 10(2):93.

[6] Unberath P, Knell C, Prokosch HU, Christoph J. Developing New Analysis Functions for a Translational Research Platform: Extending the cBioPortal for Cancer Genomics. Stud Health Technol Inform. 2019;258:46-50.

[7] HL7 International Clinical Genomics Work Group. Reporting Implementation Guide. Available at http://hl7.org/fhir/uv/genomics-reporting/index.html. Accessed January 12, 2021.

[8] Smirnov P. About FHIR facades (part I) - two approaches. Available at https://www.health-samurai.io/articles/about-fhir-facades-part-i-two-approaches. Accessed January 13, 2021.

[9] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods of Information in Medicine. 2018 jul;57(S 01):66–81.

[10] Reimer N, Busch H, Ingenerf J. Deploying a molecular tumor board on an open platform. Student Conference 2020, Medical Engineering Science, Medical Informatics, Biomedical Engineering and Auditory Technology. 2020:139-142