Using DICOM headers in FHIR Database for enrichment of the Medical Informatics Initiative's core data set

Authors: Christian Bruns, Jan Maluche, Ralf Lützkendorf, Johannes Bernarding
Data Integration Center Magdeburg, Institute of Biometry and Medical Informatics, Medical Faculty, Otto-von-Guericke University, Magdeburg

Introduction

The MII core dataset defines basic (patients, encounter, diagnoses, procedures and laboratory values) and extension modules on how data are stored and made available(1). This dataset is enriched by additional information based on incremental extraction of DICOM tags from a XNAT API(2) with the goal to a better study recruitment and the building of cohorts. These extracted data were transformed to a FHIR resource and stored in a FHIR based data ware house(3).

Methods

We developed a spring batch-based approach, defining readers, processors and writers for the data(4). The reader calls a RESTful interface to assemble information about the XNAT internal structure (subjects, projects, experiments, scans). Based on XNAT internal modification timestamps, the job determines if resources needs to be created, updated or are already up-to-date. Two timestamps can be provided by XNAT: Firstly, when an experiment was added, secondly, the time if the dataset was changed, which is relevant for the mapping and interpreted in the reading process. The processor applies a modular mapping process which maps from dicom-tags to FHIR imagingstudy resource. This resource is forwarded to the writer, which stores the resource in a FHIR server including a timestamp, according to internal XNAT data.

The whole approach is packed in a one-shot docker container, which is highly customizable through configuration files. An internal dataset of 2TB MRI-scans from 3 different scanners were used for evaluation and compared to the results of linux4health/dicom-fhir-converter(5), which maps locally stored DICOM images to FHIR resources. For speed evaluation, the hole dataset is processed and timings are taken.

Results

In the quality evaluation it couldbe proven that the only differences within the JSON files are the additional extension and within the metadata. The incremental load could speed up the load by up to 97% in comparison to a batch process.

Discussion and Conclusion

By now, only basic information about the presence of images are processed as a proof of principle. By researcher's demand, the amount of information can be easily extended through the modular implementation. The deployed process can be easily included into an existing and site individual architecture and may be used for future clinical trials.

Acknowledgement

This work was funded in part by the German Federal Ministry of Education and Research (BMBF), Funding reference number: 01ZZ1801H. The funding body did not play any role in the design of the study and collection, analysis, and interpretation of data as well as writing of the manuscript.

Bibliography

1. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018 May;57(S 1):e50–6.
2. Herrick R, Horton W, Olsen T, McKay M, Archie KA, Marcus DS. XNAT Central: Open sourcing imaging research data. NeuroImage. 2016 Jan 1;124, Part B:1093–6.
3. HL7 FHIR v4.0.1 [Internet]. Available from: https://www.hl7.org/fhir/
4. Spring Batch: Overview [Internet]. Available from: https://spring.io/projects/spring-batch#overview
5. dicom-fhir-converter [Internet]. Linux For Health; 2021. Available from: https://github.com/LinuxForHealth/dicom-fhir-converter