Expanding the OUH Clinical Data Warehouse data delivery infrastructure with OMOP CDM and OHDSI tools

Olivier Bouissou[1], Dan Johansen[2], Elisabeth Ross[1]
[1]Division of Technology and Innovation, Oslo University Hospital
[2] Research Support Services, Oslo University Hospital

Background

Oslo University Hospital (OUH) is Scandinavia's largest hospital with 1.2 million patient treatments per year. The hospital features a well-established clinical data warehouse (CDW). In 2023, OUH joined the Digital Oncology Network for Europe project (DigiONE) where a set of 40 oncology variables are harmonised to OMOP and federated across six hospitals[1].

The CDW provides (1) reports and dashboards for aggregated data[2] and (2) delivery of individual patient data to quality and research registries. We focus here on the latter and present how the existing CDW protocol and infrastructure for semi-automatic data delivery can easily be improved with the use of OHDSI tools and OMOP CDM.

Methods

CDW data reporting with extension to OMOP

The data reporting infrastructure of the CDW is implemented with Oracle technology for both ETL and building and presenting the reports and analysis. The CDW currently contains data from several clinical systems, presented to clinicians through reports and dashboards.

The ETL tools and skills used to build the reporting infrastructure were reused for delivering the DigiONE project through three workstreams: (1) sourcing the variables not already present in the CDW, (2) converting variables from the CDW into OMOP, and (3) effectuating the infrastructure to implement federated studies[3].

Data delivery infrastructure

In order to deliver patient data to clinicians, the ETL and reporting tools used by the data reporting infrastructure have been repurposed to implement a data delivery infrastructure. The infrastructure allows a CDW independent Research Support team to place data orders towards the CDW on behalf of quality and research registry owners. Research Support is the data recipient with the responsibility to dispatch the data into the specific registries.

The CDW data delivery infrastructure can be described by its main functionalities:

- Specify a cohort and the datasets to be extracted for the cohort
- Use a predetermined protocol to place an order and receive data from the CDW
- Process cohort generation and data extraction orders and deliver data to the requester

We detail below the three functionalities and how we propose to amend them with new possibilities offered by the OMOP extension. See figure 1 and 2 below for a rough representation of the as-is implementation and the changes planned for using the OMOP CDM with OHDSI tools.

Results

Cohort & dataset specification.

Cohorts and datasets are formally specified by creating two kinds of reports in the data reporting infrastructure: a cohort definition report and data extraction reports.

- A cohort definition report filters the patients based on registry-specific criteria (e.g. patients diagnosed with lung cancer during COVID-19 lockdown).
- Data extraction reports specify a subset of the available variables in the CDW reporting infrastructure. The execution of data extraction reports, when applied to a given cohort definition, generates datasets in XML for Analysis (XMLA) format.

The cohort and the extracted datasets are specified with the reporting tools, making it easy to specify new sets of variables for new cohorts without IT support, and scales up easily as the hundreds of variables available in the CDW for reporting can be used for defining a cohort or to specify ad hoc datasets. Nevertheless, the classical reporting tools are not well suited to define complex cohorts and we propose to allow for the use of ATLAS for cohort specification.

Place an order and receive data according to a predetermined protocol. Orders are placed by sending Request messages (in XML format) to a file share. Responses from CDW are sent back to the same share. Request messages ask the CDW to either execute the predefined cohort definition or execute the data extraction reports. By using ATLAS to build advanced cohort definitions, each cohort will be identified by an ATLAS cohort_ID. The cohort_id will be passed in the XML request used to place a cohort definition order, in lieu of today's cohort definition report name (see figure 1).

Each Request message, identified by a GUID, is answered by a Response message, identified with the same GUID.

The Response XML file currently contains the extracted data packed in a payload (R1 in figure 2). The protocol will need a slight alteration to specify that a slice of the OMOP CDM has to be returned from the OMOP CDM. In this case, a copy of all the data related to the patients in the cohort will be sliced out of the CDM (R2 in figure 2)

Process orders. The processing of the cohort instantiation and data extraction orders is orchestrated by the ETL tool. Request messages are pulled, parsed and executed asynchronously. Web service calls are used to execute reports (WS1) or generate cohorts (WS2). The ETL tool loads execution results into the CDW and is used to produce XML response messages. The functionalities will be amended as follows:

- The cohort instantiation order will be processed as today, but instead of making a web service call to OBIEE (WS1), a web service call to WebAPI (WS2) will trigger a cohort generation and the results will be stored directly in the CDW.
- The data extraction order can remain unchanged, allowing the delivery of any data from the CDW. If the data extraction order specifies that a slice of the OMOP CDM has to be returned, the ETL tool will send a SQL command to join the instantiated cohort to tables in the OMOP CDM and all (and only) OMOP data linked to the cohort will be extracted. In the future, this could be further developed with more advanced functionality where the OMOP data returned are from a specified period.

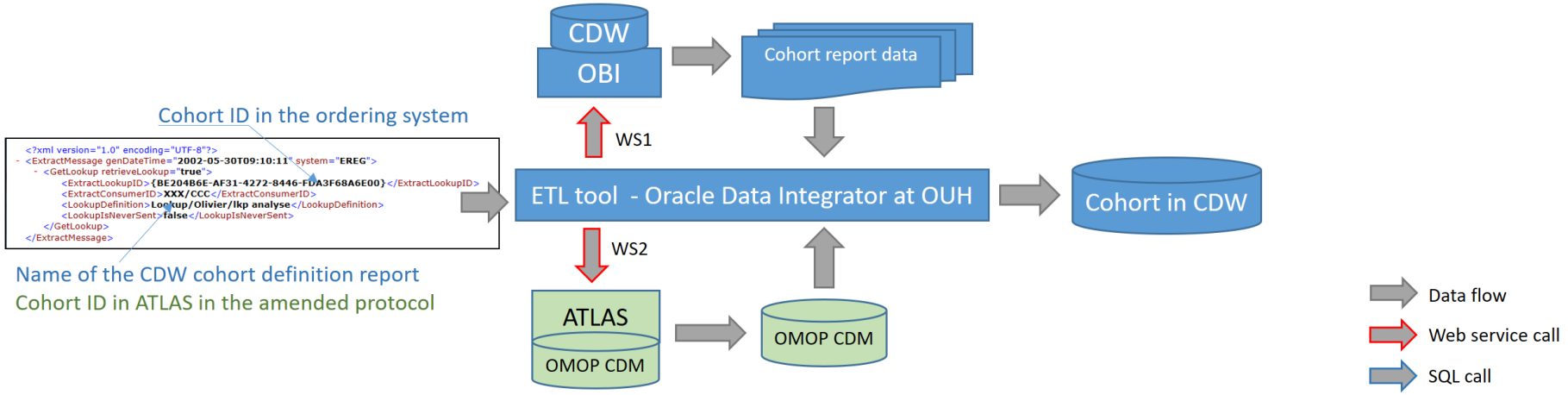

Figure 1. Parsing & execution of a cohort instantiation order. In the amended protocol, a web service call to WebAPI (WS2) triggers the generation of a cohort.

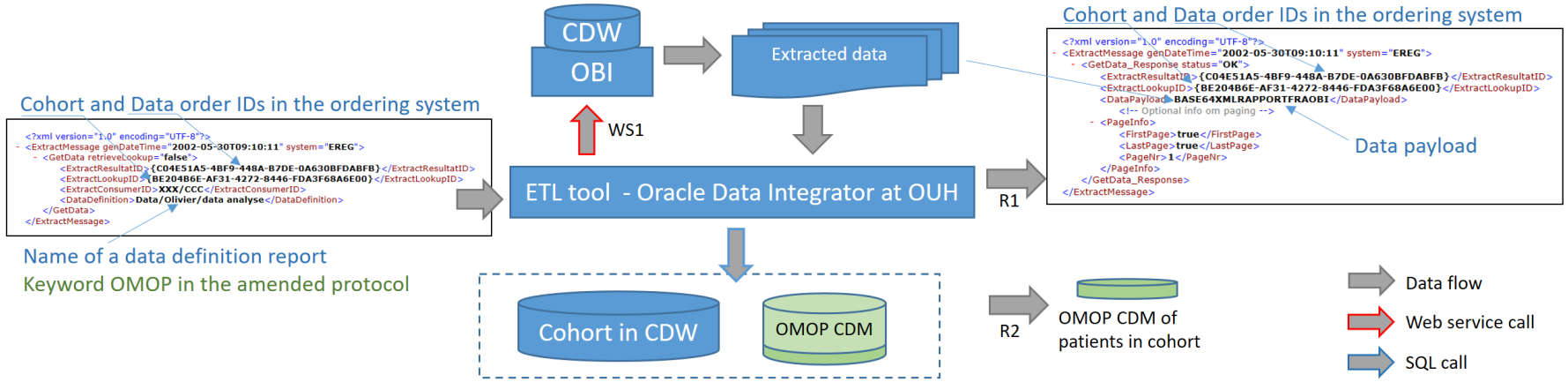

Figure 2. Parsing and execution of a data extraction order. In the amended protocol a slice of an OMOP CDM (R2) is generated instead of the result of the execution of a data definition report (R1).

Conclusion

We have shown how the CDW infrastructure for data delivery can easily benefit from the OHDSI tools and CDM for more extensive and flexible data delivery. Several extensions and further amendments to the infrastructure are considered to enable semi-automatic federated study across Europe. These new possibilities together with the ongoing work to convert gradually more data from the CDW to the OMOP CDM will support precision medicine initiatives at OUH[4].

References

1. DigiONE - Digital Infrastructure for ONcology in Europe building a European federated digital research network. DIGICORE. [Internet] 2023. https://digicore-cancer.eu/projects/5.
2. Simonsen, MCA. OUS lanserer nytt verktøy for arbeid med kreftdata. Dagens medisin. [Internet] 2022. https://www.dagensmedisin.no/kreft-oslo-universitetssykehus-ous/ous-lanserer-nytt-verktoy-for-arbeid-med-kreftdata/126897
3. Creating a sustainable framework for building federated real world evidence. OUH. [Internet] 2023. https://www.ous-research.no/files/ous-ccc/dokumenter/Bouissou%20OUS%20OMOP%20Network.pdf
4. Naderi, EH & Ross, E (2024). The Paradox of Precision Medicine. Abstract submitted for the OHDSI Europe Symposium 2024.