Oslo University Hospital Participation in a European Cancer OMOP Network

Olivier Bouissou[1], Ingrid K. S. Hanto[2], Elisabeth Ross[1]
[1] Division of Technology and Innovation, Oslo University Hospital
[2]Division of Cancer Medicine, Oslo University Hospital

Background

Oslo University Hospital (OUS) is Scandinavia's largest hospital and carries out more than 1.2 million patient treatments each year, whereof more than 12.000 new cancer patients.

The need for a comprehensive overview of cancer treatment led to a unique partnership between the Division of Cancer Medicine and the Department of Technology and eHealth, with the goal to compile cancer patient data in one dashboard in OUS' in-house Clinical Data Warehouse (CDW). A small workgroup composed of staff both from clinical and technical side was established and worked closely together to develop the Key Cancer Data Dashboard (KCDD) *(Abstract 3).*

As part of OUS' strategic move towards a structured EHR, and better use of secondary data, a new structured cancer EHR is to be deployed in 2023. This will better enable us to source and monitor structured raw data on cancer care in the CDW.

The Digital Oncology Network for Europe (DigiONE) project has been funded to get a consensus between 6 European hospitals and federate a set of 40 oncology variables describing routine clinical care and molecular information *(Abstract 1)*. The Minimal Essential Description of Cancer (MEDOC) dataset *(Abstract 2)*, will be federated into an OMOP CDM.

Participation in DigiONE fits perfectly with OUS' strategic shift towards a structured EHR and improved use of secondary data. It simultaneously gives us an opportunity to join a community that will improve the patient outcomes.

Methods

In order to minimise the delivery risk, we organised the DigiONE project internally as a continuation of the KCDD project, expanding the project organisation and the CDW infrastructure in order to build MEDOC on top of our CDW. The main benefit being that we stay within a well-functioning organisational and technical framework, where regulatory constraints, information governance and privacy are enforced.

We have adapted the CDW roadmap for 2023/2024, and the proposed delivery will roughly consist of three main workstreams (WS1, WS2 and WS3 in Figure 1):

- Workstream 1 is about sourcing DigiONE/MEDOC variables not initially present in the CDW.
- Workstream 2 is dedicated to OMOP conversion based on the curated data as soon as they are delivered into the CDW.
- Workstream 3 is effectuating the infrastructure for initiating analysis, and subsequently executing local analyses and participating in federated analyses.

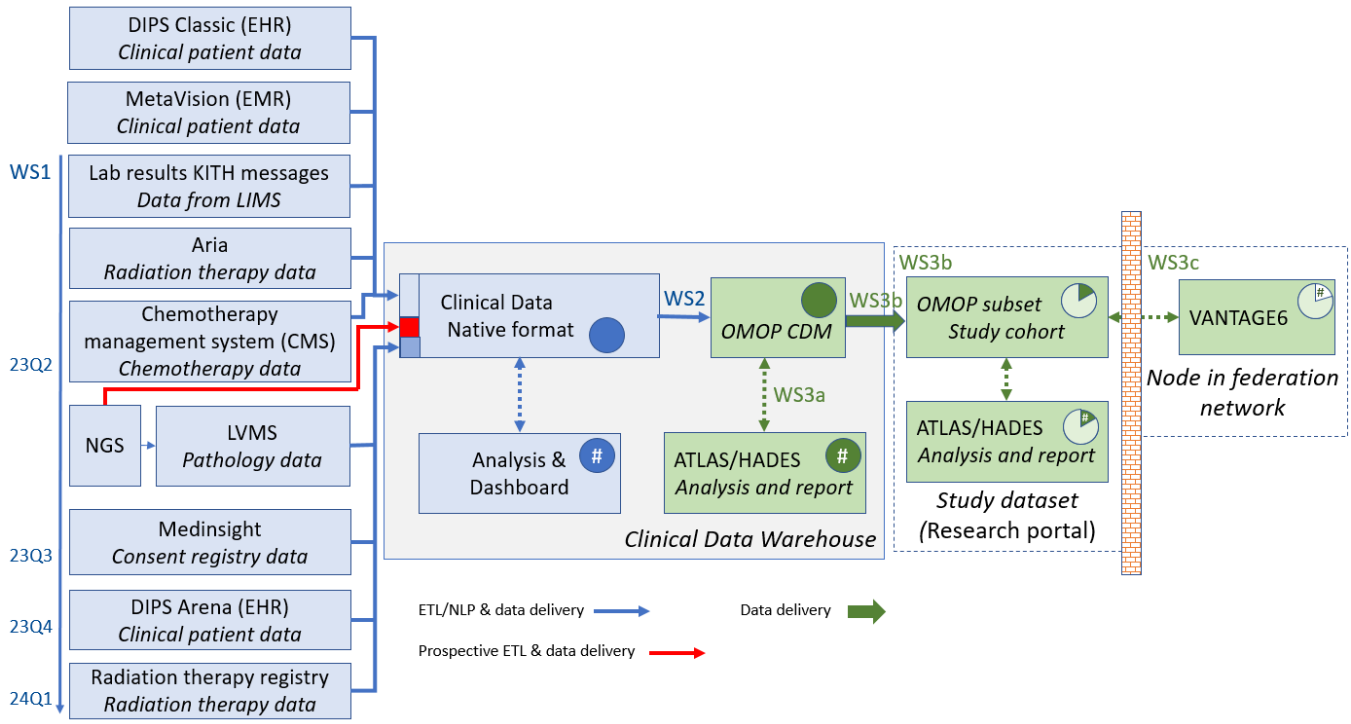

Figure 1: *Extending OUS Clinical Data Warehouse to deliver DigiONE federated studies*

Results

The three proposed workstreams can be partially parallelized, and the clear separation of concerns between the workstreams is key in risk-reducing our delivery.

Workstream 1: Source MEDOC variables into CDW

60% of the target variables are already stored in structured form in our source systems. Those data are either present in the CDW today or are on the CDW roadmap for 2023/2024. The complex variables that are registered in free text today, will be sourced from an upgraded version of the EHR system (DIPS Arena) allowing for registration of structured cancer records on prospective patients. Using the new structured cancer EHR as a source will allow us to get data quality and completeness without NLP, provided the data are registered by the clinicians. The DigiONE project ultimately delivers an RWE "clinical study ready" dataset with high quality and completeness. We expect it will be a powerful incentive for clinicians to improve primary data registration. In order to validate the quality and completeness of the data we will reuse the CDW practices and tools dedicated to that purpose.

The last category of data, biomarker and molecular data, are the most complex data to source into the CDW. In the long run we plan to use an in-house bioinformatics pipeline to populate the CDW with structured biomarker and molecular data, but initially we will resort to very simple NLP (regular expression) to extract from the pathology reports the information necessary for the studies in scope for DigiONE.

Workstream 2: Convert CDW data into OMOP CDM

The CDW infrastructure provides a sturdy foundation for organising and converting part of our CDW data model to an OMOP CDM. For converting the required data into OMOP CDM, we will use the ETL and reporting tools that are already in use in the CDW. Some of the coding schemes in use in Norway, like biochemical tests coding, are based on OUS-specific and Norwegian-specific codes. We will use USAGI to interact with clinicians and define lookup tables between our native code and OMOP CDM concepts.

Converting part of our CDW data model and concepts to an OMOP CDM will both expand our skillset and enable us to better estimate later conversions. The conversion plans have also led to participation in a newly established national OMOP network that we already benefit from and can contribute to.

Workstream 3: Effectuate the infrastructure for local and federated analyses

In order to run analyses, we will both reuse the existing CDW infrastructure for dashboarding and secure data delivery and extend our infrastructure with the OHDSI tools and a tool for federated learning.

We have identified three kinds of use cases:

- Inside the CDW context, analyses run on all patients in the MEDOC dataset in order to build a cohort or to evaluate the feasibility of a study (WS3a in Fig. 1).
- Inside the hospital context, analyses and studies are run on a subset of the patients in the MEDOC dataset (WS3b in Fig. 1). The study dataset is based on the cohort defined in WS3a. Results and statistics can be used for publication of federated results.
- As a party in a federation network, analyses are run locally on a subset of the patients in the MEDOC dataset and the results are shared through a federated learning infrastructure such as VANTAGE6[1] (WS3c in Fig. 1).

Conclusion

Participation in DigiONE gives us the possibility to reality check our ambition of a seamless provisioning of real-world datasets for federated studies. We have identified technical, organisational and legal hurdles and by crossing these the CDW team expect to build new skills and gain insight into how to best implement that vision.

Authors and Affiliations:

Author	Affiliation	Email address
Olivier Bouissou	Oslo University Hospital	Olibou@ous-hf.no
Ingrid K. S. Hanto	Oslo University Hospital	Insmal@ous-hf.no
Elisabeth Ross	Oslo University Hospital	Roseli@ous-hf.no

References

1. Moncada-Torres A, Martin F, Sieswerda M, Van Soest J, Geleijnse G, PhD. VANTAGE6: an open source priVAcy preserving federaTed leArninG infrastructurE for Secure Insight eXchange. *AMIA Annual Symposium Proceedings* 2020; 2020:870-877.

List of referenced abstracts submitted by the Digital Oncology Network for Europe (Digicore)

#	Abstract title	Abstract File name
1	Conceptual architecture for the Digital Oncology Network for Europe - an OMOP based European federated, automated cancer care quality ecosystem	'Mahon_Conceptual Architecture for Digicore_2023symposium'
2	Definition of a minimum target set of data concepts for European cancer care quality	'Mahon_Minimum data set for cancer care_2023symposium'
3	Developing a Key Cancer Data Dashboard at Oslo University Hospital	'Hanto_OUS_KeyCancerDashboard_2023symposium'