Building precision oncology via real-world data standardization, privacy preserving infrastructures, and collaboration among IT and clinical experts: The DIGICORE network experience

A.Traverso[1,2], G. Pahlavan[2], X. Fernandez[2], P. Mahon[2], G. Tonon[1,2]
[1]Vita-Salute San Raffele University, Milan, Italy
[2] Digital Institute For Cancer Outcomes Research (DIGICORE), Brussels, Belgium

RATIONALE

DIGICORE: Advancing Precision Oncology with Real-World Data

- ❑ **Unlocking RWD for Precision Oncology:** Real-World Data (RWD) is crucial for personalized cancer treatments, offering insights beyond clinical trials and enabling international comparisons of treatment effectiveness.
- ❑ **Addressing RWD Challenges**: Significant hurdles exist in using RWD, including data sparsity, unstructured formats, privacy regulations, and varied IT infrastructure.
- ❑ **DIGICORE's Solutions**: The DIGICORE consortium tackles these challenges through initiatives like standardized data elements (MEDOC), AI tools for data digitization, privacy-preserving federated learning, and clinician training, with multiple projects already underway or planned.

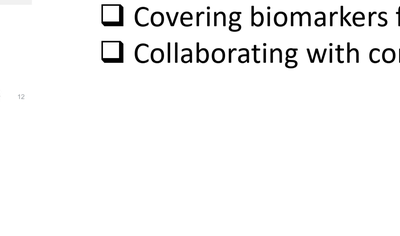
THE NETWORK

DIGICORE is an international public private partnership set-up to transform and digitise cancer outcome research, and from there digitise research

*Representative sample

THE MISSION

DIGIOCORE has developed an open innovation technology stack to solve this, called DigiONE – a digital Oncology Network for Europe


https://doi.org/10.1038/s41591-023-02715-8

A federated learning system for precision oncology in Europe: DigiONE

Piers Mahon, Ismini Chatzitheofilou, Andre Dekker, Xosé Fernández, Geoff Hall, Aslaug Helland, Alberto Traverso, Cedric Van Marcke, Janne Vehreschild, Gennaro Ciliberto & Giovanni Tonon

Check for updates

DigiONE is a pilot European learning health system in precision oncology that aims to identify optimal cancer treatments by learning from every patient, not just those in trials, through privacy-preserving interrogation of their standardized routine electronic health records.

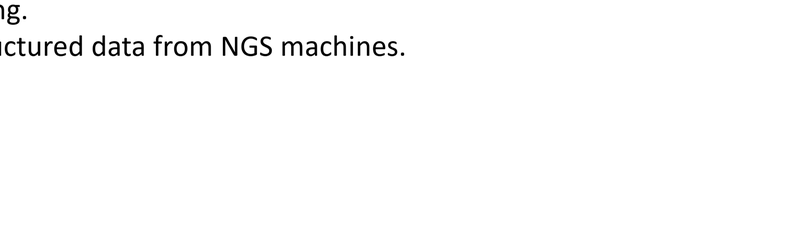
Every cancer is different, depending on its particular molecular subtype. Cancer treatments will only be effective on certain subtypes. The optimal matching between patients and treatments will require gram-

explore new indications for approved targeted and immunotherapies.

1. **Minimal Essential Description Of Cancer** (MEDOC)
2. **International semantic interoperability from new Cancer OMOP standards**
3. **Pan-format Cancer data ingestion**. Not just ETL also NLP, OCR
4. **GDPR recital 34 privacy conserving solutions for Next Generation Sequencing**
5. **Full federation with open source Vantage6** to allow statistical analysis equivalent to centralised data, but without data pooling
6. **Modular, protocolized implementation plans** to solve for limited data normalisation skills in most hospitals
7. **All in open standards and vendor agnostic**

THE DATA

Those hospital DigiONE databases cover all cancers or >70K new diagnosis a year, typically with cohort start dates of around 2018



THE PROJECTS

10 DIGICORE clinical protocols are in delivery or development, and our first technology research project – we'll hear from 4 of them today

Level	1. Digital Research Basics	2. Structured data in OMOP	3. High quality MEDOC (in a given cancer)
Technical Approach	• Raw electronic health records • Manual curation and data extraction possible on small cohorts to a protocol specific common data model	• Local database built from structured data, typically rich in demographics and activity / treatments • But weak clinical phenotypes	• Local database with high data completeness on all MEDOC items, likely using NLP for clinical phenotype and biomarker items
Example DIGICORE research projects	1. *Head & Neck* 2. *Prostate* 3. *Breast – dose variation* 4. *CRC*	5. *Covid19* 6. *Proposed disease burden and Time-to-treatment study*	7. *DINASTY mNSCLC (Francesca)* 8. *DINASTY HEr2- mBreast* 9. *Planned CRC DINASTY* 10. *Planned gender program*

11. Impact of different NLP systems on data quality

NEXT STEPS

WORKING GROUPS: Performing Effective RWE research

- ❑ Committees within DIGICORE to connect like-minded RWE researchers in hospitals.
- ❑ Aligned to major research themes - e.g. a tumour type or topic like health inequalities.
- ❑ Open for rectruitment to all European researchers.

TRAINING PROGRAMS: encourage broader study protocol participation

- ❑ On-going: IDEAL4OMOP, IDEAL4RWE.
- ❑ Proposed: Federated learning introduction, OMOP study hackatons, gender medicine, grant writing.

OMOP: Extending coverage to -omics

- ❑ Covering biomarkers from Next Generation Sequencing.
- ❑ Collaborating with companies to directly produce structured data from NGS machines.