European Network of Cancer Registries

ENCR Workshop on Software and AI Tools for CRS
Oncological clinical minimum data set and its mapping to the OMOP Common Data Model

Paolo Lasalvia

Evaluative Epidemiology Unit
Department of Epidemiology and Data Science
Fondazione IRCCS Istituto Nazionale dei Tumori di Milano
paolo.lasalvia@istitutotumori.mi.it

25 October 2024

FONDAZIONE IRCCS
ISTITUTO NAZIONALE
DEI TUMORI

European Commission

EURACAN registry

1. Sarcomas
2. Head & neck cancers
3. Rare thoracic cancers
4. Digestive rare cancers
5. Neuroendocrine tumours
6. Endocrine gland tumours
7. Central nervous system tumours
8. Rare female genital cancers
9. Rare urological and male genital tumours
10. Rare skin cancers & non-cutaneous melanoma

Minimum dataset

- ✓ To help describe the natural history

- ✓ To evaluate factors that influence prognosis (e.g. mortality, survival, progression free survival and treatment response)

- ✓ To assess treatments effectiveness (systemic, radiotherapy, surgery, target therapy, immunotherapy and possible combinations)

- ✓ To measure indicators of quality of care (diagnostic and staging procedures, treatment strategies, follow-up etc.)
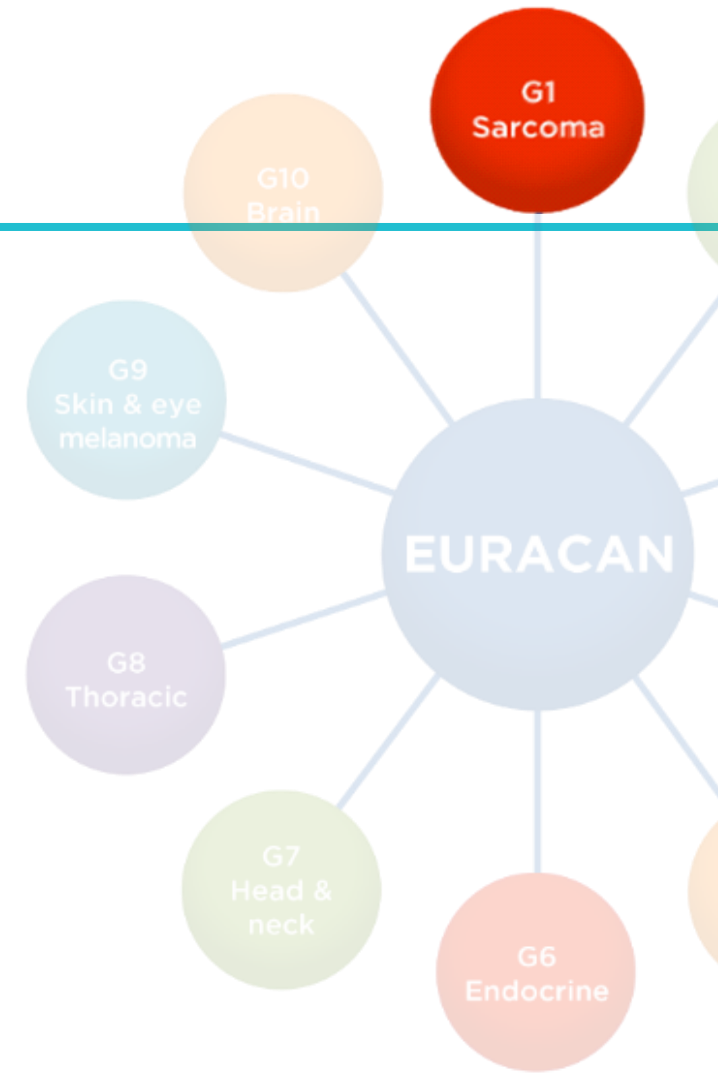
IDEA4RC

IDEA4RC Goal

To establish a **Data Space for rare cancers** that will make possible the secondary use of existing multisource health data (cancer registry data, national registries, data from biobanks etc.) across EU healthcare systems leveraging emerging interoperability technologies and AI approaches.

Observational Medical Outcomes Partnership (OMOP) Oncology Common Data Model

Open-source process

- Join the collaborative

- Propose a study to the open collaborative

- Write protocol

  http://www.ohdsi.org/web/wiki/doku.php?id=research:studies

- Code it, run it locally, debug it (minimize other's work)

- Publish it: https://github.com/ohdsi

- Each node voluntarily executes on their CDM

- Centrally share results

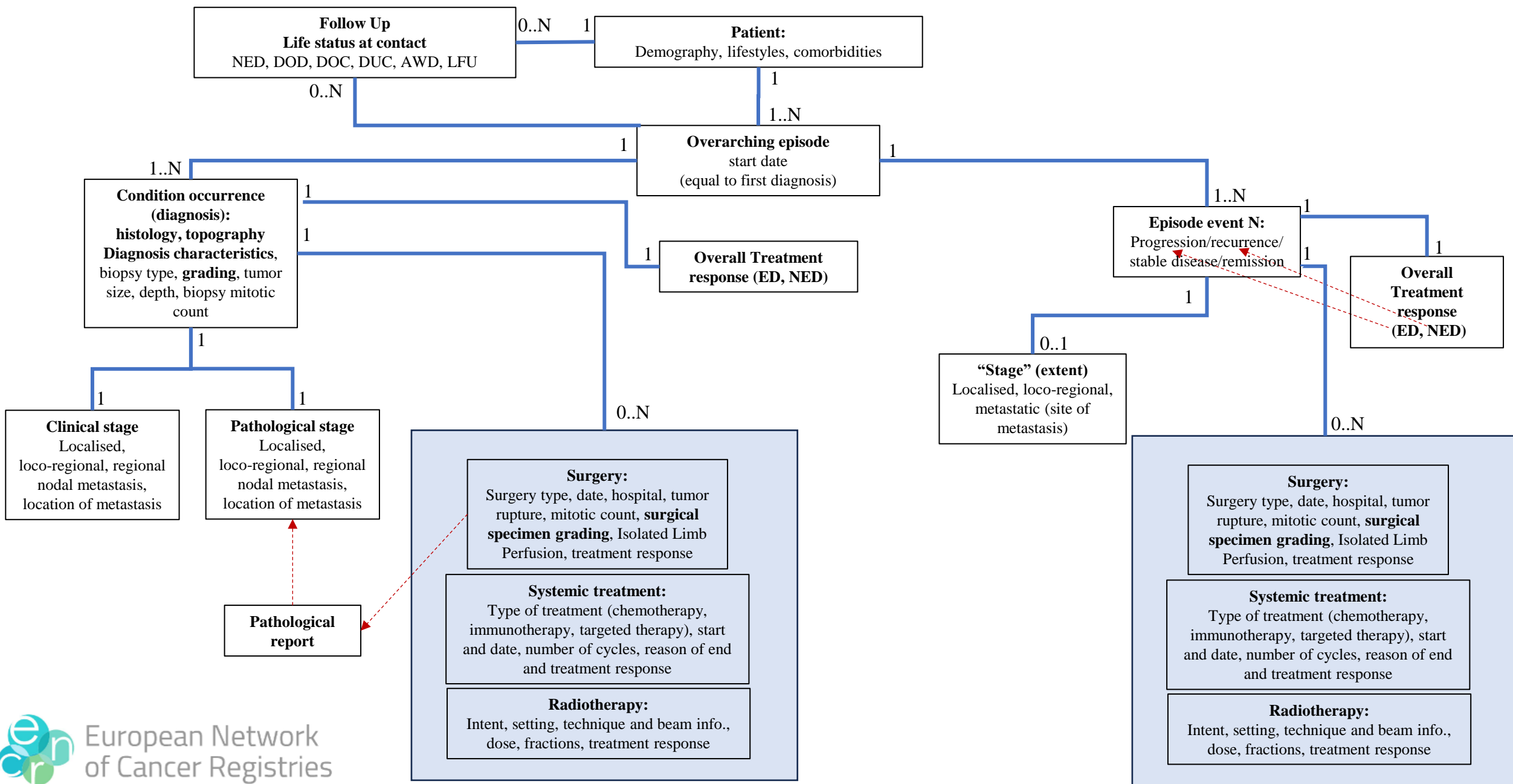- Collaboratively explore results and jointly publish findings

IDEA4RC model data elements

Minimum data set

- **Patient:** demography, lifestyles, comorbidity, previous cancer

- **Overarching episode:**
  - CONDITION OCCURRENCE (diagnosis) and cancer diagnosis characteristics (histology, topography, biopsy type, grading, tumor size, depth, biopsy mitotic count)
  - CLINICAL STAGE, PATHOLOGICAL STAGE (Localised, loco-regional, regional nodal metastasis, location of metastasis)
  - CANCER TREATMENT (surgery, systemic treatment, radiotherapy) with TREATMENT RESPONSE
  - EPISODE EVENT (Progression/Recurrence/Stable disease/Remission)

- **Patient Follow up** with Life status at last contact

European Network of Cancer Registries

European Commission

IDEA4RC Entity and Relationships Model

European Network of Cancer Registries
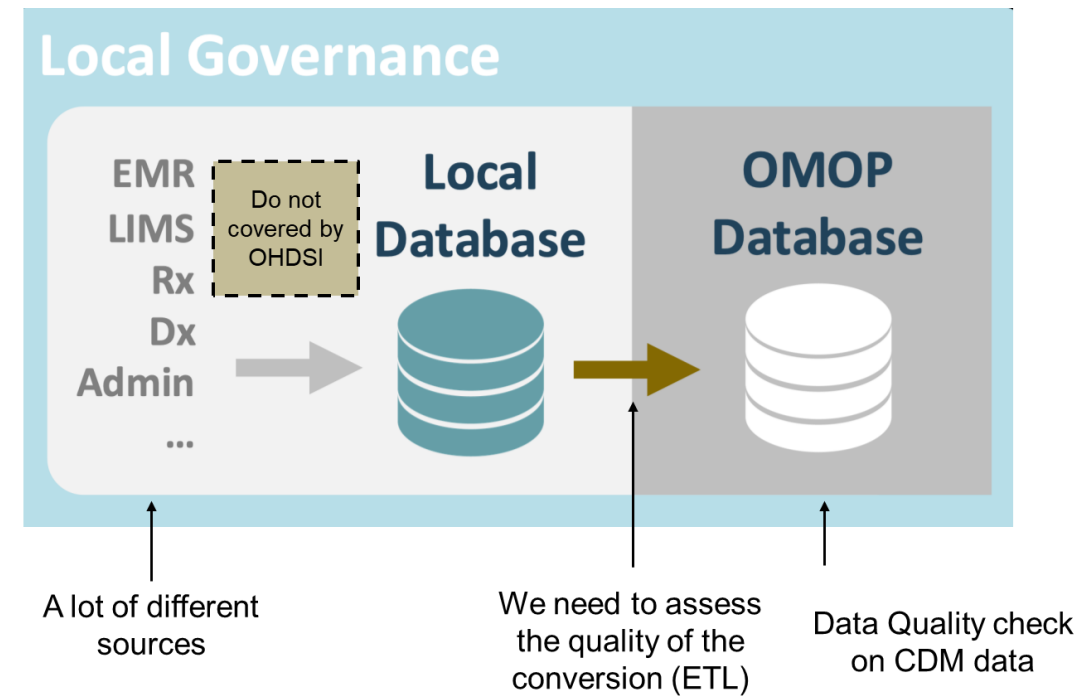
Minimum dataset vs OMOP

- Every variable of the minimum data set is mapped to OMOP

- Only a <u>few differences</u> in the IDEA4RC model: treatment is slightly different to enhance interoperability

- The OMOP oncology model is a <u>work in progress</u>, the OHDSI Oncology Workgroup is refining all the terminology and updating Vocabularies

Tips and tricks with OMOP mapping
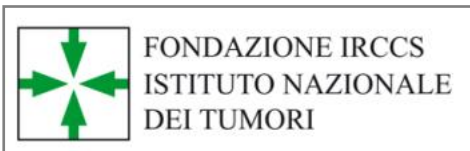
- We need a <u>repository</u> to collect concept codes not present at the moment in OMOP Vocabularies

- <u>Custom concepts</u> to input our data in the model

- <u>Data quality assessment</u> of both the source database and the OMOP one, to validate the quality of the ETL (Extract, Transform and Load) process

- Study specific data quality checks

Local Governance

EMR
LIMS
Rx
Dx
Admin
...

Do not covered by OHDSI

Local Database

OMOP Database

A lot of different sources

We need to assess the quality of the conversion (ETL)

Data Quality check on CDM data

www.encr.eu

Thank you for your attention!

FONDAZIONE IRCCS
ISTITUTO NAZIONALE
DEI TUMORI

Paolo Lasalvia
paolo.lasalvia@istitutotumori.mi.it

European Network
of Cancer Registries

European
Commission