Intelligent ecosystem to improve the governance, the sharing, and the re-use of health data for rare cancers

Deliverable 10.3

Standards for a Rare Cancer Data Ecosystem

26 February 2025

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101057048

Distribution List

Organization	Name of recipients
1 - Coord INT	A. Trama, P. Casali, L. Buratti, P. Baili, J. Fleming, L. Licitra, E. Martinelli, G. Scoazec
2 - UDEU	A. Almeida, , U. Zulaika Zurimendi, N.. Kalocsay
3 - MME	F. Mercalli, S. Copelli, M. Vitali
4 - UPM	E. Gaeta, G. Fico, L. Lopez, I. Alonso, C. Vera, A. Estevan, V. G. Dominguez, I. Alonso, L. Hernandez, C. Vera
5 - HL7	G. Cangioli, C. Chronaki, R.Gazzarata
6 - ECCP	S. Ziegler, A. Quesada, S. Schiffner, V. Tsiompanidou
7 - ENG	A. Sperlea, E. Mancuso, M. Melideo, F. Saccà, V. Falanga, M. Rosa
8 - CERTH	K. Votis, A. Triantafyllidis, N. Laloumis
9 - UU	S. van Hees, Wouter Boon, E. Moors, M. Kahn-Parker, C. Egher
10 - DICOR	C. Lombardo, G. Pesce, G. Ciliberto, G. Tonon,
10.1 - ACC (Affiliated)	L. Villanova, L. Villanova, P. De Paoli, G. Piaggio, N. Sulemane, M. Pallocca, A. De Nicolo.
11 - FBK	A. Lavelli, S. Poggianella, O. Mayora, A.M. Dallaserra
12 – IKNL	E. Bosma. G. Geleijnse, A. Van Gestel
13 – CLB	A. Sans, M. Brahmi, A. Pons, J-Y Blay, H. Crochet, J. Olaz, J. Bollard, C. Chemin-Airiau, H. Crochet
14 - APHP	B. Baujat, E. Koffi
15 - FJD	J Martin-Broto, N. Hindi, M. Martin Ruiz, A. Montero Manso, C. Roldàn Mogìo, D. Da Silva, A. Herrero, B. Barrios
16 - VGR	Magnus Kjellberg, L. De Verier, A. Muth
17 – MSCI	I. Lugowska, D. Kielczewska, M.Rosinska, A KAwecki , A., P. Rutkowski
18 - MUH	R. Knopp, A. Sediva, K. Kopeckova, A. Nohejlova Medkova, M. Vorisek
19 - OUS	J. Nygård, M. Sending, O. Zaikova
20 - MMCI	J. Halamkova, I. Mladenkova, l. Tomastik, V. Novacek, T. Kazda, I. Mladenkova, O. Sapožnikov
21 - CLN	R. Szmuc, J. Poleszczuk, R. Lugowski
22 - FPNS	M. Barbeito Gomez, P. Parente, L. Carrajo Garcia, P. Ramos Vieiro
23 - TNO	E. Lazovik, L. Zilverberg, S. Dalmolen
24 - INF	ML Clementi, C. Sabelli
25 - UKE	S. Bauer, S. Lang, S. Mattheis, N. Midtank
26 - NIPH	S. Larønningen,

Revision History

Revision	Date of Issue	Author(s)	Brief Description of Change
1.0	2025-02-28	R. Gazzarata, C. Chronaki, G. Cangioli (HL7)	First release
2.0	2025-02-25	L. Villanova, G. Piaggio, N. Sulemane (ACC), G. Ciliberto, G. Tonon (DICOR)	Peer review
3.0	2025-02-26	R. Gazzarata	Final release to integrate comments, suggestions and corrections by the peer reviewers
3.0	2025-02-28	R. Gazzarata	Approval by Coordinator

Addressees of this document

This document is addressed to the whole IDEA4RC Consortium. It is an official deliverable for the project and shall be delivered to the European Commission and appointed experts.

TABLE OF CONTENTS

LIST OF FIGURES

`

Abbreviations and definitions

Abbreviation	Definition
1+MG	1+Million Genomes
AI	Artificial Intelligence
AI4HI	Artificial Intelligence for Health Imaging
API	Application Program Interface
B1MG	Beyond 1 Million Genomes
CAMP FHIR	Clinical Asset Mapping Program for FHIR
CMD	Common Data Model
CS-AIW	Cancer Survivorship - AI for Well-being
DIGIONE	DIGItal Oncology Network for Europe
EBCP	Europe's Beating Cancer Plan
ECIS	European Cancer Information System
EEHRxF	European EHR Exchange Format
EFMI	European Federation of Medical Informatics
EHDS	European Health Data Space
EHR	Electronic Health Record
ENCR	European Network of Cancer registry
ePS	Electronic Patient Summary
ERN(s)	European Reference Network(s)
ETL	Extract, Transform, Load
FAIR	Findable, Accessible, Interoperable, and Reusable
FHIR	Fast Healthcare Interoperability Resources
GDI	Genomic Data Infrastructure
HDS	Health Data Sharing Platforms
HDW	Health Data Warehouses
IRCCS	Istituto di Ricovero e Cura a Carattere Scientifico (Scientific Hospitalization and Treatment Institute)
IG	Implementation Guide
IKNL	Netherlands Comprehensive Cancer Organisation
IMI2	Innovative Medicines Initiatives 2
IRC	Internet Relay Chat
JRC	Joint Research Center
MDX	Molecular Diagnostic Information
MEDOC	Minimal Essential Description of Cancer
MIE	Medical Informatics Europe
MII	Medical Informatics Initiative
NCATS	National Center for Advancing Translational Sciences
NLP	Natural Language Processing
NKR	Dutch Cancer Registry
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
ONC	Office of the National Coordinator for Health Information Technology
OSTP	Office of Science and Technology Policy
PHOENIX	Patient Health Oncological Expertise Network for International Exchange
PS	Patient Summary
RC	Rare Cancer

RD	Rare Disease
REST	REpresentational State Transfer
RWE	Real-Word Evidence
RWD	Real-World Data
SDO	Standards Developing Organization
STD	Standard for Trial Use
SurPass	Survivorship Passport
TI	Terminology Infrastructure
URL	Uniform Resource Locator
WG	Working Group
WGM	Working Group Meeting

1 EXECUTIVE SUMMARY

This deliverable is part of the Task T10.3 "Ecosystem standards" and reports on standards that are essential for the enlarged ecosystem, and the activities performed to analyse the context and to embark a journey to harmonize the ecosystem standards. This involved continuous monitoring of the relevant landscape emerging from each core technical development work package (WP3, WP4, WP5, WP6, WP7) to ensure all tools and infrastructure components developed within the project adhere to relevant existing standards and ongoing standardization initiatives performed also by other European projects or national initiatives.

This deliverable addresses the following goals. First, to support an effective ecosystem enlargement from a technological point of view and to ensure that all tools and infrastructure components are available through OpenAPI. Finally, to address and incorporate emerging European and international standards ensuring the highest degree of interoperability among various data sources and repositories.

This document presents the activities performed from month 13 to month 30 of IDEA4RC project on the analysis of standards for a Rare Cancer Data Ecosystem and reports on standards relevant for the enlarged ecosystem achieved as result of Task 10.3. Finally, the documents report the activities performed in IDEA4RC project to start to build a European Cancer Common Data Model.

In the first period, the activities have been focused on the collaboration and the active participation in some initiatives and Working Groups promoted by HL7 Int (e.g. Codex and Vulcan FHIR Accelerator) or jointly promoted by the HL7 Int and other communities such as OHDSI (e.g. "OHDSI OMOP + FHIR" Working Group). Next, an HL7 Europe Working Group (WG) was created to prepare this deliverable and to start working on the creation of a European Cancer Common Data Model. A collaboration was established with other European funded projects (also considering the 2 project clusters in which IDEA4RC is involved, which are InToEHR and EU-funded projects on data-driven approaches in cancer) and with national initiatives, all focused on the cancer domain.
The WG first met during the "HL7 Europe Working Group Meeting and HL7 FHIR Marathon"[1] in Athens in January 2024 at the "European Cancer Mission Track". During this event a first survey

[1] https://hl7.eu/wgm2024/

phase was performed with the participants of the event. Following meetings were taken place online to examine the visibility on the European Cancer Common Data Model to support primary and secondary use of data, based on HL7 FHIR and OMOP aligned with other relevant cancer data models as mCODE, MEDOC and OSIRIS. A working paper facilitated the collaboration and the agreement.

A workshop was organized in August 2024 as part of MIE 2024 conference in Athens to ask support to collaborate with relevant EFMI groups. After this event a second survey phase was performed involving also other European funded projects and asking more relevant information. The results of this survey were used as input to prepare the "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025"[2], that was hosted in Lisbon in February 2025.

In the next chapter, an overview of the European and international context on cancer data ecosystem is described. Then, the methodology adopted, and the results obtained are presented. Finally, the document reports the conclusions discussing the achieved results and the next steps.

[2] https://hl7europe.eu/hl7-working-group-meeting-2025/

2 INTRODUCTION

In this section the current European and international context on cancer data ecosystem is presented. In detail, first the European context, with the Europe's Beating Cancer plan, and some EU-funded and national projects, and with the international network of cancer research centres and their corresponding data models or datasets, are presented. Finally, a global overview on the ongoing HL7 International initiatives as Vulcan (FHIR to OMOP), Codex (mCODE) is provided.

2.1 The European context: project and initiatives

Considering the significant impact of cancer in Europe, in terms of incidence and economic costs, the need for updated strategies in cancer prevention, treatment, and care, especially considering the COVID-19 pandemic's disruptions, was recognized. In 2021 the European Union (EU) has responded to these needs with the "Europe's Beating Cancer Plan"[3] (EBCP) aimed to address cancer comprehensively through 4 pillars: prevention, early detection, treatment, and improving patients' and survivors' quality of life. It will help Member States turn the tide against cancer, preparing 10 flagship initiatives and 32 supporting actions planned between 2021 and 2030.

One of the EBCP objectives is to facilitate the development and to employ new technologies, research and innovation for patient-centred cancer prevention and care. To achieve this goal EBCP works closely with the Horizon Europe Mission on Cancer, which drives EU investment in cancer research and innovation. These two initiatives will provide a better understanding of cancer, allow for earlier diagnosis and optimisation of treatment and improve cancer patients' quality of life during and beyond their cancer treatment. With a €4 billion budget, the EU Cancer Plan makes use of several available funding instruments (e.g. the EU4Health, Horizon Europe, Digital Europe programmes). IDEA4RC represents one of the projects funded by the Horizon 2020 programme.

The Europe's Beating Cancer Plan seeks to make the most of the potential of data and digitalisation. For this reason, it wants to exploit the potentialities of the European Health Data

[3] https://health.ec.europa.eu/system/files/2022-02/eu_cancer-plan_en_0.pdf

Space (EHDS) to enable cancer patients to securely access and share their health data in an integrated format in the electronic health records (EHRs) between healthcare providers and across borders in the EU. The EHDS was announced by the EC in May 2022 as "*a health-specific data sharing framework establishing clear rules, common standards and practices, infrastructures and a governance framework for the use of electronic health data by patients and for research, innovation, policy making, patient safety, statistics or regulatory purposes.* The EHDS was recently provisionally accepted by the European parliament and council in March 2024. The EDHS scope is twofold; on one hand to improve the primary use of data, thanks to the electronic cross-border health services provided by MyHealth@EU, and on the other to enable the re-use of data (secondary use), thanks to the cross-border infrastructure HealthData@EU. To achieve these goals, the EHDS proposal establishes a new regulatory framework for manufacturers of EHR systems, including compliance with new European interoperability specifications, the European EHR Exchange Format (EEHRxF) with the objective of effective use of health data.

Figure 1 European EHR Exchange Format (EEHRxF) logo

2.1.1 The European projects

In this context, different European projects were funded by the European Commission to develop tools and/or infrastructures to advance cancer research by supporting secondary data use with different specific purposes. With Horizon Europe Programme, IDEA4RC, FLUTE, AIDAVA, canSERV, EOSC4cancer, I3LUNG, and ONCOVALUE were funded, while with Horizon 2020 Programme, PanCareSurPass, INCISIVE, ProCAncer-I, and DECIDER were funded. The Digital Europe Programme funded EUCAIM and GDI, and the Innovative Medicines Initiatives 2 (IMI2) founded OPTIMA, and PIONEER. Finally, with the EU4Health Programme, EHDS2Pilot, BEACON were funded.

Figure 2 X-eHealth project logo

The X-eHealth[4] is a Horizon 2020 project promoted to lay the groundwork for a practical, interoperable, secure and cross-border EEHRxF, working towards the improvement of the eHealth sector. It also addressed the specific needs in terms of data elements and semantic standards for rare diseases in general and for rare cancers as far as they differ from rare diseases, in relation to well-defined use cases for both planned and unplanned care. Data elements were defined based both on previously agreed minimum data sets for rare disease (RD) patient registration, which are in some cases implemented in health records in some countries, and on the experience of European Reference Networks (ERNs) - please see 2.1.3 – on rare cancers (RC).

Figure 3 FLUTE project logo

Focusing on prostate cancer, the FLUTE[5] project works on novel ways of using data through a privacy-preserving approach. The project aims to improve predictions of aggressive prostate cancer through AI support to physicians, while minimizing unnecessary biopsies, ultimately benefiting patients and reducing associated costs. To train the AI, a platform will be developed to access the data while preserving the necessary privacy.

[4] https://www.x-ehealth.eu/
[5] https://www.fluteproject.eu/

Figure 4 AIDAVA project logo

The AIDAVA[6] project will develop a digital solution, orchestrating diverse artificial intelligence technologies, for more efficient curation and publishing of personal health data, delivering interoperable and reusable personal health records for the benefit of patients and clinical researchers.

Figure 5 canSERV project logo

The canSERV[7] project provides cutting edge, interdisciplinary and customised oncology services across the entire cancer continuum. The aim is to offer a comprehensive portfolio of oncology-related research services available to all scientists in EU member countries, associated countries and beyond

Figure 6 EOSC4cancer project logo

The EOSC4Cancer[8] project will make diverse types of cancer data accessible: genomics, imaging, medical, clinical, environmental and socio-economic. It will use and enhance

[6] https://aidava.eu/
[7] https://www.canserv.eu/
[8] https://eosc4cancer.eu/

federated and interoperable systems for securely identifying, sharing, processing and reusing FAIR data across borders and offer them via community-driven analysis environments

Figure 7 I3LUNG project logo

The I3LUNG[9] project aims to prove better assistance and individualize treatment for patients affected by metastatic lung cancer.

Figure 8 ONCOVALUE project logo

The ONCOVALUE[10] project will work to implement value-based oncology care to help improve cancer care by enabling and guiding cancer clinics to collect, harmonize and analyze high-quality Real-World Data (RWD) in real-time and by developing an AI-based framework for assessing real-life effectiveness of novel cancer therapies in real-time.

Figure 9 PanCareSurPass project logo

[9] https://i3lung.eu/
[10] https://oncovalue.org/

The PanCareSurPass[11] project aims to improve person-centered care for childhood and adolescent cancer survivors by integrating the digital tool Survivorship Passport (SurPass) into the Electronic Health Information Systems of clinics in Austria, Belgium, Germany, Italy, Lithuania, and Spain.

Figure 10 INCISIVE project logo

The INCISIVE[12] project aims to develop and validate an AI-based toolbox that enhances the accuracy, specificity, sensitivity, interpretability and cost-effectiveness of existing cancer imaging methods. One of INCISIVE's goals is to develop an Interoperable pan-European federated repository of clinical data and medical images that allows sharing data in compliance with legal, ethical, privacy and security requirements, for AI-related training and experimentation. INCISIVE clinical data was semantically encoded in a way that gets everyone within a system to speak the same language and understand the meaning of the data. Templates contain input fields about clinical elements and laboratory elements; it was distinguished between these two types of data and used SNOMED CT or LOINC according to the medical concepts. With DICOM and HL7 FHIR, INCISIVE achieves a Common Data Model (CDM) that allows receiving, storing, and processing information from multiple sources, multiple data providers, as a one standard way. For more detail, see INCISIVE Standardization Suggestions, and the HL7 FHIR INCISIVE Implementation Guide.

Figure 11 ProCAncer-I project logo

[11] https://www.pancaresurpass.eu/
[12] https://incisive-project.eu/

The ProCAncer-I[13] project proposes to develop advanced artificial intelligence models to address unmet clinical needs: diagnosis, metastases detection and prediction of response to treatment. To achieve this, partners will generate a large interoperable repository of health images, and a scalable high-performance computing platform hosting the largest collection of prostate cancer Magnetic Resonance Images used for developing robust prostate cancer AI models.

Figure 12 DEDICER project logo

The DECIDER[14] project wants to develop diagnostic tools and treatments for high-grade serous ovarian cancer with the help of AI methods. The aim is to identify earlier those patients who do not respond well to the first-line treatments, and to find effective treatments for patients with a drug-resistant cancer. It also studies the legal issues that impede or slow down the use of new treatments to facilitate commercialization and availability of personalized therapies in an ethically and legally sustainable manner.

Figure 13 EUCAIM project logo

The EUCAIM[15] project aimed to establish a pan-European digital federated infrastructure called Cancer Image Europe. This platform gradually aggregates FAIR (Findable, Accessible, Interoperable, and Reusable) cancer-related images from real-world data and existing repositories from 5 large EU-funded projects on big data and AI in cancer imaging, while preserving data sovereignty. The initiative includes an Atlas of Cancer Images and seeks to

[13] https://www.procancer-i.eu/
[14] https://www.deciderproject.eu/
[15] https://cancerimage.eu/

facilitate the development and benchmarking of AI tools for Precision Medicine, as well as federated data nodes and federated AI-tools deployment. The project targets clinicians, researchers, and innovators, offering a foundation for validated clinical decision-making systems in areas such as diagnosis, treatment, and predictive medicine. By collaborating with clinical data providers, researchers, research infrastructures, and industry, the project addresses the challenges of implementing a comprehensive cancer imaging infrastructure. It aims to overcome fragmentation in existing repositories by leveraging Artificial Intelligence for Health Imaging (AI4HI) initiative repositories, European research infrastructures, and national/regional repositories, encompassing clinical images, pathology, molecular, and laboratory data. The project aligns with the EHDS initiative to establish a sustainable flagship federated repository for high-quality data and tools.

Figure 14 Genomic Data Infrastructure project logo

The Genomic Data Infrastructure (GDI)[16] project will enable access to genomic and related phenotypic and clinical data across Europe. It will be possible by establishing a federated, sustainable and secure infrastructure to access the data. It builds on the outputs of the Beyond 1 Million Genomes (B1MG)[17] project and aim to implement the goals put forward by the 1+Million Genomes (1+MG) initiative. B1MG was a European project to help in the creation of a network of genetic and clinical data across Europe. This initiative was a commitment of 24 EU countries, the UK and Norway to give cross-border access to one million sequenced genomes by 2022.

Figure 15 OPTIMA project logo

[16] https://gdi.onemilliongenomes.eu/
[17] https://b1mg-project.eu/

The OPTIMA[18] want to implement a vast federated and centralised network of European data providers to help answer the highest priority research questions in prostate, breast and lung cancer, especially where the existing evidence underpinning clinical practice guidelines is weak or lacking. In parallel, the development and application in clinical settings of comprehensive dynamic computer-interpretable guidelines to better support shared decision making by clinicians and patients.

Figure 16 PIONEER project logo

The PIONEER[19] project aims to transform the field of prostate cancer care with particular focus on improving prostate-cancer related outcomes, health system efficiency and the quality of health and social care across Europe.

Figure 17 The HealthData@EU Pilot logo

In October 2022, the EHDS2Pilot[20] project started, bringing together 17 partners including health data access bodies, health data sharing infrastructures and European agencies. It was a two-year long European project that built a pilot version of the EHDS infrastructure for the secondary use of health data, named "HealthData@EU" as previously mentioned. The project connected data platforms in a network infrastructure and develop services supporting the user journey for research projects using health data from various EU Member States. It also provided guidelines for data standards, data quality, data security and data transfer to support this cross-

[18] https://www.optima-oncology.eu/
[19] https://prostate-pioneer.eu/
[20] https://ehds2pilot.eu/

border infrastructure. Priority services include a metadata discovery service and a common health data access request. The consortium collaborated closely with the European Commission and their team working on developing the central services for secondary use of health data.

Figure 18 Cancer Care BEACON project logo

The Cancer Care BEACON[21] project is a groundbreaking initiative aimed at enhancing cancer care across the European Union by mapping cancer centers' capacities and compiling oncology datasets, among other cancer-related resources, and presenting them in an interactive, accessible application for all stakeholders. Designed to support patients, researchers, policymakers, and healthcare providers, the BEACON Decision-Support Application offers a comprehensive platform for informed decision-making and facilitates transatlantic collaborations and clinical trials.

Figure 19 The Cancer Survivorship - AI for Well-being logo

[21] https://wiki.beaconcancer.org/index.php?title=Cancer_Beacon

Finally, the Cancer Survivorship - AI for Well-being (CS-AIW)[22] is a cluster which brings together European-funded projects working in Artificial Intelligence (AI) for healthcare and well-being, primarily in the post-cancer treatment space, with a collective aim to cross-fertilize and transcend individual project experiences for the wider good and outcome of their individual projects. A holistic view of the cancer survivor is considered in all these projects, thus leading to the collection of not only clinical information but as well data gathered at the patient's environment (e.g. ePROMS and sensor data) and can be integrated at the cancer survivor's smart card. A typical data model implementation that has been proposed and could be a starting point for integrating self-reported and lifestyle data is namely CASIDE[23], which is based on HL7 FHIR.

2.1.2 National initiatives

In Europe there are different national initiatives related to cancer: OSIRIS, ONCOFAIR, German Cancer Registries / German Medical Informatics Initiative (MII) Oncology Module, the Netherlands Comprehensive Cancer Organisation (IKNL) and the Italian Health Big Data project.

Figure 20 The OSIRIS logo

OSIRIS[24] is a national project launched by the Integrated Cancer Research Sites (SIRICs, French comprehensive cancer centers) accredited by the French National Cancer Institute. The network has proposed a list of 130 clinical and genomic items. This “OSIRIS set” establishes a minimum dataset for the sharing of clinic-biological data in oncology. The list is based on a conceptual and temporal cancer disease model agreed upon within the group.

ONCOFAIR[25] is a French national project aimed to define and test a methodological and technical framework for upstream/downstream interoperability of Health Data Warehouses

[22] https://www.cs-aiw.eu/
[23] González-Castro, L., Cal-González, V. M., Del Fiol, G., & López-Nores, M. (2021). CASIDE: a data model for interoperable cancer survivorship information based on FHIR. Journal of biomedical informatics, 124, 103953
[24] https://en.e-cancer.fr/OSIRIS-a-national-data-sharing-project
[25] https://www.kereval.com/nouveau-projet-dinteroperabilite-des-donnees-oncologiques/

(HDW) / Health Data Sharing (HDS) platforms by taking the case of use of chemotherapy data. Developed by Kereval and the LTSI Laboratory in Rennes, the aim is to make this data easy to find, accessible, interoperable and reusable in the logic of the FAIR DATA paradigm. The ultimate objective is to harmonize methodologies both nationally and across Europe, ensuring interoperability on a broader scale.

Figure 21 The German Medical Informatics Initiative logo

The German Cancer Registries / German Medical Informatics Initiative (MII) Oncology Module[26][27] are a federated system of 15 state cancer registries. It developed from the collaboration between the more clinically-focussed ADT register with the more epidemiological GEKID registry and has been continuously collecting data since 2013, with regular updates to the core dataset. The current dataset is called oBDS, short for oncologic core data set, and is sent to via a defined XML data. The Medical Informatics Initiative was founded in 2016 in a collaborative effort from all German University Hospitals to make data from routine care available digitally, reliably and quickly for medical research. For this purpose, a national FHIR-based Core Dataset was developed, currently providing more than 14 different modules. As of today, data from hospitals is made available by internal Data Integration centres and can be queried using the Research Data Portal Health (Forschungsdatenportal Gesundheit, FDPG). The latest addition is the Extension Module Oncology, which is based on the oBDS data model, and enables the integration of cancer-specific data points. In addition, the MII is working on an Extension Module for Molecular Tumor Boards, covering information in the area of precision medicine, integrating structured pathological and genomical information.

[26] https://www.medizininformatik-initiative.de/en/about-initiative
[27] https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/k/krebsregister

Figure 22 *The* The Netherlands Comprehensive Cancer Organisation logo

The Netherlands Comprehensive Cancer Organisation (IKNL) is an independent knowledge institute for oncological and palliative care. IKNL's mission is to reduce the impact of cancer basing on insights from real world data in the Dutch Cancer Registry (NKR). IKNL makes an important contribution to reducing the impact of cancer by collecting and making essential and reliable data available and offering information and insights to healthcare professionals, policy makers and researchers.

Figure 23 The Italian Health Big Data project logo

The Italian Health Big Data project, coordinated by the IDEA4RC partner ACC, aims to develop an integrated and federated Cloud Platform for the collection, sharing and advanced analysis (AI) of real-time clinical-scientific data of hundreds of thousands of patients from Italian research hospitals (IRCCS). The fine goal is to create a universal knowledge resource in medicine, based on respect for patient privacy and free access for doctors, scientists and patients and to improving the diagnosis and treatment of each individual patient by analysing the data of all patients, and seeking new treatments.

2.1.3 European networks and research infrastructures

The European Reference Networks (ERNs) are virtual patient-centred networks involving healthcare providers and patient representatives across Europe. Working together, these stakeholders aim to tackle complex or rare diseases and conditions that require highly specialized care and concentrated knowledge and resources.

Figure 24 The EURECAN logo

EURACAN[28] is the ERN for all rare adult solid cancers. In 2019, EURACAN initiated an hospital-based registry focusing on 2 out of the rare cancers family namely head and neck cancers and sarcomas. It is a federated registry leveraging the federated learning as a privacy preserving solution. Core data elements include the "Set of common data elements for Rare Disease Registration" developed by Joint Research Center (JRC) to address specificities of rare cancers as compared to rare diseases. Furthermore, following the EURACAN registry objectives, data are collected on patient characteristics, exposure and outcomes. Patient characteristics will consist of descriptive patient data, such as information on patient demographics, including race, lifestyle, medical history, health status, etc. Exposure data will focus on the disease, devices, procedures, treatments or services of interest. Outcome data will describe patient outcomes (e.g. survival, progression, progression-free survival, death, etc.). In addition, data on potential confounders will also be collected. The registry is active for the rare head and neck cancers, for soft tissue sarcomas and for ultra-rare sarcomas i.e. the epithelioid hemangioendothelioma. The soft tissue sarcoma registry has been mapped with the OMOP oncological model.

Figure 25 The European Network of Cancer Registries logo

The European Network of Cancer registry (ENCR)[29] is a network of population-based cancer registries existing since the 70's in Europe. In 2012 the Joint Research Center (JRC) of the EU started to act as the secretariat of the ENCR and to coordinate the centralisation of data from cancer registries to develop the ECIS, the European Cancer Information System. There is call

[28] https://euracan.eu/
[29] https://www.encr.eu/call-for-data

for data based on an agreed protocol and standards, data quality checks software and data quality control run at centralised level. The ENCR-JRC[30] project was launched in 2015 by the ENCR Steering Committee and JRC to set up a standardized and comparable database for monitoring cancer incidence and mortality in the European Union and to provide regular information on the burden of cancer in Europe.

Figure 26 The BBMRI-ERIC logo

BBMRI-ERIC[31] is a European research infrastructure for biobanking and biomolecular resources to bring together all the main players from the biobanking field – researchers, biobankers, industry, and patients – to boost biomedical research. To that end, it offers quality management services, support with ethical, legal and societal issues, and a number of online tools and software solutions with the final goal to make new treatments possible. It was involved in EHDS2Pilot as pilot.

Figure 27 The DiGICORE logo

DIGICORE[32], one of the IDEA4RC partners, is a pan-European research network built to accelerate the implementation of precision oncology in Europe. DIGICORE promotes and equips cancer centres in their use of routine EHR and molecular diagnostic information (MDX) for trial automation, real world outcomes research, digital diagnostics and care quality management. The ultimate goal is to shape a digital research infrastructure based on digital interoperability between its Members. Network membership supports them to improve data

[30] https://www.encr.eu/encr-jrc-project
[31] https://www.bbmri-eric.eu/
[32] https://digicore-cancer.eu

quality and completeness, develop new data sources and tools, share digital best practices and promote novel, digitally enabled research methods. The objectives are:

- Enhancing cancer outcomes in Europe and drive innovation through the building of a federated digital research infrastructure to perform protocolised observational studies, appropriate medical analytics or screening tools and support pragmatic trials.
- Focus on gathering high-quality, up-to-date real-world data on cancer diagnosis, treatment, outcomes, and other related conditions.
- Creating digital interoperability between its Members and Associate Members using high quality multisite real-world data.
- Develop standards for files and network protocols supporting them to improve data quality and completeness; new data sources and tools; novel digital enabled research methods (especially in biomarker development).

2.1.4 DIGIONE and MEDOC

The DIGItal Oncology Network for Europe (DIGIONE)[33] is a European project coordinated by DIGICORE and funded by HADEA which aims to create a federated digital research network that links routine, high quality clinical data with routine molecular data information from >10 large cancer centres in several countries. The underlying digital infrastructure provides a minimal description of every patient's cancer diagnosis, biomarkers, treatment and outcomes in near real time relying on a Minimal Essential Description of Cancer (MEDOC)[34] built upon consensus and aligned with international standards. It provides a minimal description of cancer from diagnosis to outcome, and includes all major research inclusion/exclusion criteria, to create a unique resource for high-quality real-word evidence (RWE) and care quality management. An open innovation programme leveraging interoperability technologies (OMOP) improving primary data capture with Natural Language Processing (NLP) solutions. New research services will follow with high quality structured real-world data (RWD) from routine cancer care with privacy-preserving data analytics (federated AI) addressing some urgent clinical research questions that require scale.

MEDOC was created to support the automation of outcomes research based on DIGICORE's collective experience in international real-world outcomes research studies. The intent was to

[33] https://digicore-cancer.eu/Allegati/DIGIONE_Press_release_22_12.pdf
[34] Mahon, Piers, et al. "A federated learning system for precision oncology in Europe: DigiONE." Nature Medicine (2024): 1-4.

use the research experience and "OSIRIS minimum data set for data sharing and interoperability in oncology"[35] to create the smallest ever minimal cancer dataset by selecting key clinical concepts that describe cancer sufficiently for outcomes research, are clinically important and have reasonable EHR availability across Europe. The resulting key elements defined in MEDOC are organized in 5 categories: demographics, clinical phenotype, biomarkers, treatment and outcomes (Table 1).

MEDOC Categories	MEDOC concepts
1. Demographics	1.1 Date of birth (month)
	1.2 Sex
	1.3 Weight (with timestamp)
	1.4 Height
	1.5 Healthcare ID (or other unique identifier)
	1.6 Legal basis for data processing
2. Clinical phenotype	2.1 Primary cancer diagnosis and comorbidities, typically in International Classification of Disease standards such as ICD10, ICD9 or ICD-O-3
	2.2 Charlson comorbidity index (derived from 17 comorbidities in 2.1)
	2.3 Date of primary cancer diagnosis
	2.4 Method of primary cancer diagnosis
	2.5 Performance status (for example, coded by ECOG or Karnofsky standards)
	2.6 Disease stage in a recognized standard such as TNM
	2.7 Histological cell type, typically in ICD-O-3 standards
	2.8 Menopausal status (for example, for patients with breast cancer)
3. Biomarkers	3.1 Biomarker name
	3.2 Biomarker measure
	3.3 Biological sample ID
4. Treatment	4.1 Line of therapy (derived algorithmically within each cancer type)

[35] Guerin, Julien, et al. "OSIRIS: a minimum data set for data sharing and interoperability in oncology." JCO Clinical Cancer Informatics 5 (2021): 256-265.

	4.2 Anti-cancer treatment name, including systemic treatment and supportive therapy
	4.3 Molecule generic name
	4.4 Start date for drug treatment
	4.5 Treatment dose
	4.6 End date for drug treatment
	4.7 Radiotherapy type
	4.8 Radiotherapy start date
	4.9 Radiotherapy dose
	4.10 Radiotherapy end date
	4.11 Surgery type
	4.12 Surgery date
	4.13 Participation in clinical trial
	4.14 Date of trial consent
5. Outcomes	5.1 Date of death, at any location
	5.2 Time to next treatment (derived from treatment start dates)
	5.3 Metastasis presence/absence
	5.4 Metastasis location
	5.5 Date of clinical visits (with cancer related visits separated from other visits)
	5.6 Vital status (derived from visits or death linkage)
	5.7 Extent of debulking surgery (for example, for patients with gynecological cancer)

Table 1 Data concepts defined in MEDOC

2.2 HL7 Int initiatives: Vulcan (FHIR to OMOP), Codex (mCODE)

In the USA more than 2 millions new cancer cases and more than 600,000 cancer deaths were predicted for 2024[36]. For this reason, the White House Office of Science and Technology Policy (OSTP) is leading an effort to enhance the clinical trial infrastructure in the USA, aiming to make trials faster, more inclusive, and efficient and, in this context, it is supporting the Biden Cancer

[36] https://cancerx.health/

Moonshot goal to eradicate cancer[37]. The OSTP is working with the Department of Health and Human Services' Office of the National Coordinator for Health Information Technology (ONC) and the HL7 FHIR Accelerator community[38] on faster clinical trial data capture.

Figure 28 The HL7 FHIR Accelerator logo

The HL7 FHIR ACCELERATOR program is designed to support communities and collaborative groups across the global healthcare spectrum in creating and adopting high-quality FHIR Implementation Guides, promoting health data interoperability. FHIR Accelerators operate within the HL7 organization but are separate initiatives from regular HL7 International initiatives. The vision of HL7 FHIR is based on the rapid global acceptance of FHIR standards as an innovative platform for data interoperability. HL7 acts as a facilitator in standards development, helping communities use FHIR to address common use cases.

Vulcan[39], together with CodeX[40] and other HL7 FHIR Accelerators, is stepping up to speed the development of standardized approaches to data exchange, with the goal of piloting faster and more inclusive data capture for multi-site clinical trials.

Figure 29 The Vulcan FHIR Accelerator logo

[37] https://www.whitehouse.gov/ostp/news-updates/2023/10/26/a-stronger-clinical-trial-infrastructure-for-better-health-outcomes/
[38] https://www.hl7.org/about/fhir-accelerator/
[39] https://confluence.hl7.org/display/VA
[40] https://www.hl7.org/codex/

In particular, Vulcan FHIR Accelerator was born from an idea of a group of invested representatives from government agencies, academia, technology companies, standards development organizations, patients, and industry consortiums to connect clinical research and healthcare. Vulcan will bring together stakeholders across the translational and clinical research community to bridge existing gaps between clinical care and clinical research, strategically connect industry collaboratives, maximize collective resources, and deliver integrated tools and resources, through the adoption of HL7 FHIR standard to support the bidirectional flow of data. The "FHIR to OMOP" is one of the Vulcan FHIR Accelerator projects[41], created to support the development of FHIR to OMOP data transfer for better analysis of clinical data for research. Considering that the OMOP CDM (Comon Data Model) developed by OHDSI community is widely adopted for generating high-value evidence based on observational data while HL7 FHIR it's the world's most important interoperability platform for health data. The project aims to establish a stable set of transformations between FHIR and OMOP to permit broader utilization of observational data in research systems, enabling FHIR API access to OMOP databases and data on FHIR servers to populate OMOP instances. This in-turn supports generation of transformed data that are comparable and consistent, increases the reliability of data, generation of knowledge artifacts, and reproducible research results. As result of this "OHDSI OMOP + FHIR" Working Group the FHIR to OMOP FHIR IG[42] is under development to reduce implementation costs, increase the speed of ETL in projects and increase the quality of transformed data for a core set of patient data. This creates the foundation for consistent automation of OMOP + FHIR data transformations.

The "OHDSI OMOP + FHIR" Working Group is formed by 4 Project Subgroups[43]: Data Model Harmonization, OMOR + FHIR Terminologies, FHIR-OMOP Digital Quality Measurement Use Case and FHOR-OMOP Oncology Use Case.

Figure 30 The CodeX FHIR Accelerator logo

[41] https://www.hl7vulcan.org/
[42] https://build.fhir.org/ig/HL7/fhir-omop-ig/
[43] https://confluence.hl7.org/pages/viewpage.action?pageId=79506058

CodeX is a Member-driven HL7 FHIR Accelerator hosting a growing community working together to enable FHIR-based interoperability that drives substantial improvements around the most important challenges and opportunities in patient health. In the Oncology space, CodeX Members are achieving interoperability by integrating and testing the mCODE (minimal Common Oncology Data Elements) FHIR Implementation Guide[44] - an open standard language for cancer data - within Use Cases that test new workflows supporting better cancer care and research. mCODE® is a trademark owned by the American Society of Clinical Oncology, but unfortunately mCODE FHIR Implementation Guide is not a global specification, but a US specific one, requiring a dependency with the US Core (https://www.hl7.org/fhir/us/core/) rules and it is focused on the primary data use. For these two reasons, it was adopted only in some parts of the IDEA4RC FHIR Implementation Guide as indicated in the Deliverable D3.1 "FHIR Implementation Guide". However, considering this importance it was considered as a starting point for the creation of the first draft of the European Cancer Common Data Model as it will be reported in chapter 3.2.5**Errore. L'origine riferimento non è stata trovata.**.

[44] https://hl7.org/fhir/us/mcode/index.html

3 METHODOLOGY

In this section the methodology adopted is presented, starting from the one adopted in the first period of activities based on the collaboration and participation to the HL7 Int initiatives Vulcan (FHIR to OMOP), and Codex (mCODE), and followed by the one used in the activities performed at European level with the HL7 Europe Working Group "European Cancer Mission".

3.1 Collaboration and participation to Vulcan (FHIR to OMOP) and Codex (mCODE)

The very first months have been dedicated to the preparatory activities for establishing liaisons with Standards Developing Organizations (SDOs) and collecting inputs from technical IDEA4RC WPs. Based on the WP2 and WP5 feedback, the focus has been initially put on the cooperative use of OMOP and HL7 FHIR. For this reason, HL7 Europe established a personal connection with the participants to the "OHDSI OMOP + FHIR" Working Group of the Vulcan FHIR Accelerator "FHIR to OMOP" project, collecting artefacts and tools used, during the HL7 International Working Group Meetings (WGMs) in January 2023 which were hosted in Henderson (USA).

The HL7 Int WGMs are held three times per year at varying locations and serve two important purposes:

1. To provide HL7 International work groups a chance to meet face to face to work on the standards as well as the opportunity to network with industry leaders from around the world.

2. To provide an invaluable educational resource for the healthcare IT community.

More than 40 HL7 Working Groups are dedicated to specialized areas of interest and are directly responsible for the content of the standards and spend much of their time at the WGMs hard at work on standards development. Attending a WGM can be a great way to keep up-to-date on what is happening in a particular area, and everyone attending an HL7 WGM is invited to join any of the working groups.

All the materials collected by HL7 Europe during the HL7 Int WGMs in January 2023 have been shared with the appropriate IDEA4RC WPs and follow up contacts have been realized with the "OHDSI OMOP + FHIR" Working Group based on the WP3 needs of identifying OMOP artifacts (e.g. Athena concepts) in HL7 FHIR to discuss specific issues and topics. For some specific discussions, it was necessary to interact also with other HL7 and OHDSI WG as the HL7 Int Terminology Infrastructure (TI) WG[45] and the OHDSI CDM Vocabulary Subgroup[46].

During the activities the collaboration with the overall FHIR community was also fundamental. In fact, HL7 FHIR consists of two parts:

- The FHIR specifications – the core specification[47] as well as an extensive set of guides and other specifications that build on it.
- The community of individuals who engage with these specifications, build solutions that leverage them, and work to improve them.

HL7 invites and encourages all to engage the community by adopting some free selected tools such as chat.fhir.org, HL7's Confluence, and HL7's Jira.

chat.fhir.org is the HL7 FHIR instance of Zulip that is an open-source chat and collaborative software created in 2014 where communication occurs in streams (which are like channels in Internet Relay Chat (IRC)). Each stream can have several topics – Zulip features a unique threading model, in which each message also has a topic, along with the content. Zulip claims that this improves productivity by "making it easy to catch up after a day of meetings". Apart from this, Zulip offers standard features found in collaboration apps like message reactions, message search history, polls, private messaging, group messaging etc. Zulip streams can be private or public – only people invited to a private stream can view messages in it, while anyone within an organization can join a public stream. Messages in Zulip can be sent in plain-text or formatted using markdown, along with images, links, and file attachments. Thanks to chat.fhir.org the FHIR community has a forum with a wide number of streams related to different topics in healthcare (Figure 31). The users can Subscribe to the ones of interest, set up notifications for keywords of interest, and search for pre-existing answers to questions from others and if they don't find the needed information, can create new conversation to ask specific questions. This tool was adopted during the project to investigate with the community

[45] https://confluence.hl7.org/display/VOC
[46] https://ohdsi.org/workgroups/
[47] https://hl7.org/fhir/

FHIR community on the availability for reuse of any representation on OMOP CDM as FHIR Logical Model in the specific channel "OMOP + FHIR"[48].

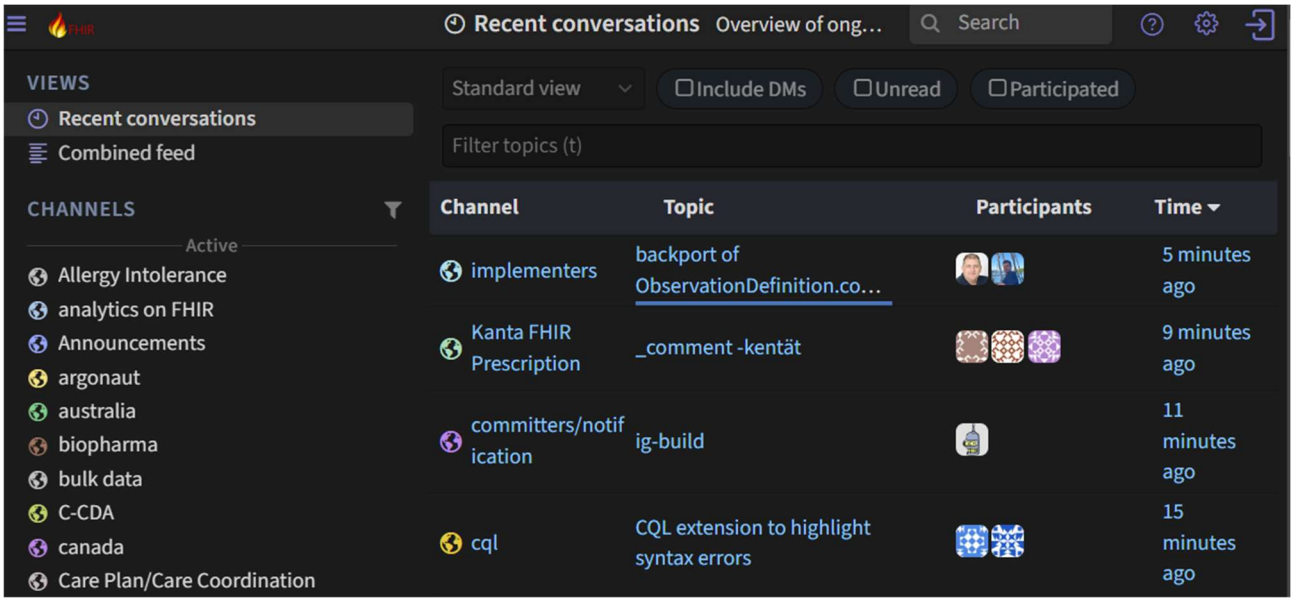


Figure 31 The HL7 FHIR community forum chat.fhir.org

Another important collaboration during the activity period was with Codex FHIR Accelerator on mCODE FHIR Implementation Guide[49]. As previously mentioned, this IG was adopted only in some parts of the IDEA4RC FHIR Implementation Guide (see the Deliverable D3.1 "FHIR Implementation Guide" for more details) but in this adoption it was necessary to ask some changes to the mCODE FHIR IG based on the IDEAR4RC feedback. For this activity, another important tool, HL7's Jira, was adopted by HL7 Europe. In fact, in case users find issues with the specification, they can submit feedback using the "Propose a change" link in the footer of any page in the specification, as shown in Figure 32. This link will take the users to HL7's Jira site (Figure 33).

Figure 32 The footer of HL7 mCODE FHIR IG

[48] https://chat.fhir.org/#narrow/stream/286658-omop-.2B-fhir/topic/OMOP.20CDM.205.2E4.20as.20FHIR.20LM/near/347066009
[49] https://hl7.org/fhir/us/mcode/index.html

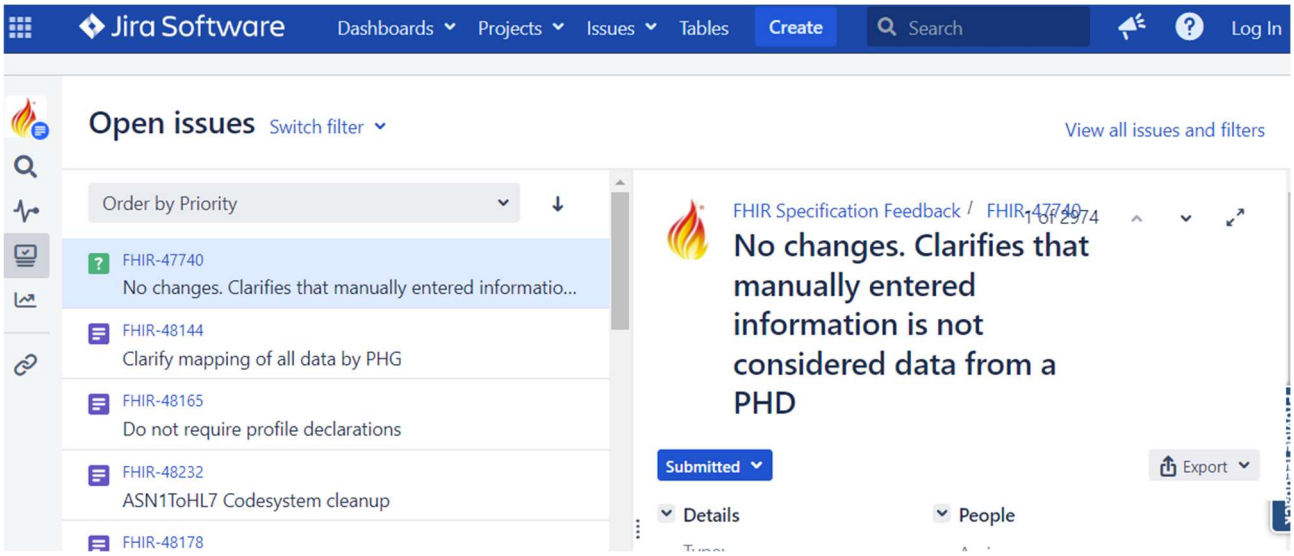


Figure 33 "Open issue" page on HL7's Jira

HL7's Jira[50] is the HL7 instance of Jira that is a proprietary product developed by Atlassian from 2002 that allows bug tracking, issue tracking and agile project management. In detail, HL7's Jira allows to manage requests for change to HL7 specifications and terminologies, as well as to manage organizational processes. While most content is publicly available, an account is necessary to access certain search functionality, to comment on existing requests, or to submit new issues. User accounts are available to anyone and does not require being an HL7 member. HL7 Jira accounts also provide access to HL7's Confluence site. HL7's Confluence[51] is the HL7 instance of Confluence that is a web-based corporate wiki developed by Atlassian from 2004 and is the HL7 HL7's community workspace to share documentation of how HL7 creates standards, the decision making records and notes across all of hL7 sub-groups, and a trove of resources about the HL7 community, its processes, and events. An example of HL7's Confluence page is "Specification Feedback" that provides a guide to submit new feedback and an overview on the feedback process.

[50] https://jira.hl7.org/secure/Dashboard.jspa
[51] https://confluence.hl7.org/

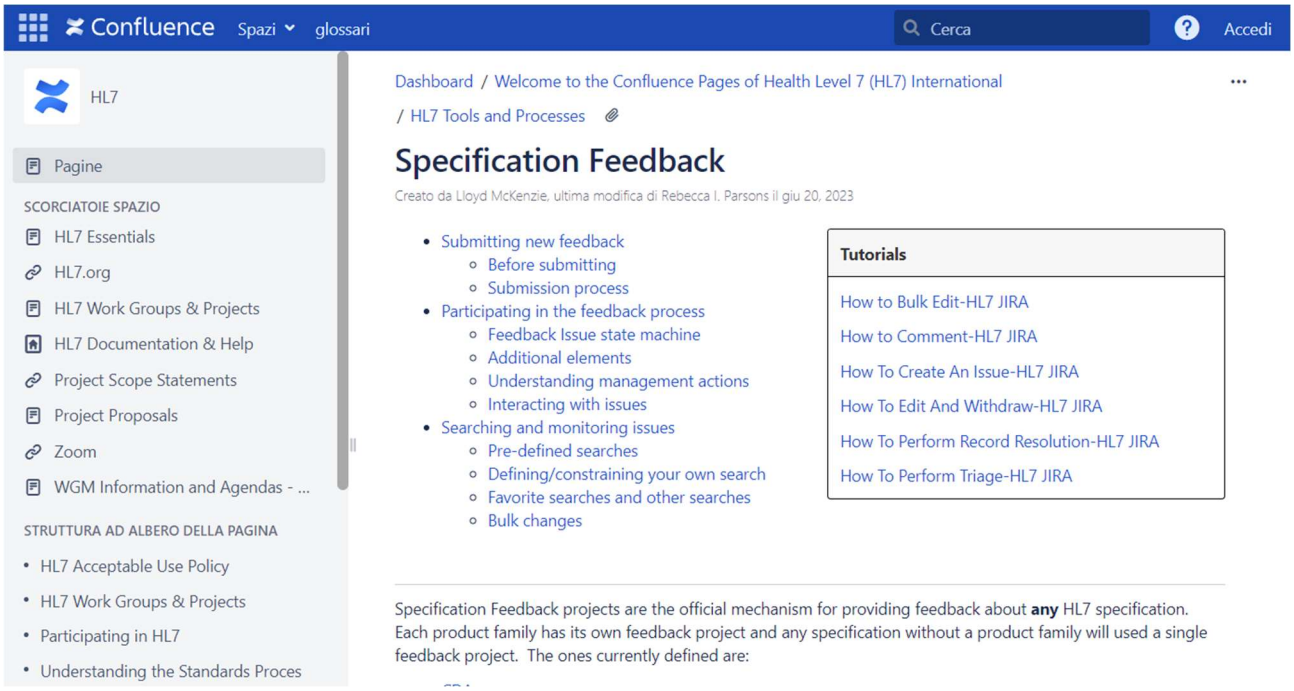


Figure 34 “Specification Feedback” page on HL7's Confluence

3.2 The HL7 Europe Working Group “European Cancer Mission”

After the activities reported in the previous chapters, at the end of 2023, taking into account the international context presented in chapter 2 and the feedbacks collected during the Working Group Meeting of the 2 project cluster in which IDEA4RC is involved (i.e. InToEHR and EU-funded projects on data-driven approaches in cancer) to search for synergies with other projects in the clusters, HL7 Europe, in accord with the IDEA4RC coordinators, decided to put its attention on cancer domain to try to improve the primary and secondary data use in the context of the EHDS.

3.2.1 Participation to HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024

For this reason, HL7 Europe proposed to the HL7 Community to lead the “European Cancer Mission Track” for the first “HL7 Europe Working Group (WG) Meeting and HL7 FHIR Marathon[52]” that was hosted in Athens in January 2024.

To officially propose the track, some calls with possible participants were organized and a specific HL7’s Confluence page was created and completed by HL7 Europe.

[52] https://hl7.eu/wgm2024/

Figure 35 The "HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024" poster

The proposed intent of this WG was to compare existing European FHIR Cancer-related scenarios and Implementation Guides (IGs) to identified gaps and achieve some levels of alignment and make an inventory of topics of interest for this domain with the final aim to support the development of EEHRxF, by creating an HL7 Europe Cancer Common FHIR Implementation Guide for cancer domain and explore new EHDS priority domains. In detail, the intent of this FHIR IG will be like US mCODE FHIR IG that is to increase interoperability by assembling a core set of structured data elements for oncology EHRs but applicable in the European context. As indicated in the Deliverable D3.1 "FHIR Implementation Guide" some profiles and elements of the IDEA4RC FHIR IG were derived from the corresponding mCODE but only where applicable.

As will be later presented in detail, during the event in Athens HL7 Europe, in accord with the IDEA4RC coordinators, decided to extend the intent to start the activity by also working on the data model to try to achieve a European Cancer Common Data Model, a fundamental step towards the creation of a HL7 Europe FHIR Implementation Guide for cancer domain.

After the approval of the proposed track, before the HL7 Europe Working Group Meeting and HL7 FHIR Marathon in January 2024 a questionnaire[53] was prepared and proposed to the participating organizations and projects to collect some information before the start of the "European Cancer Mission Track" (Figure 36).

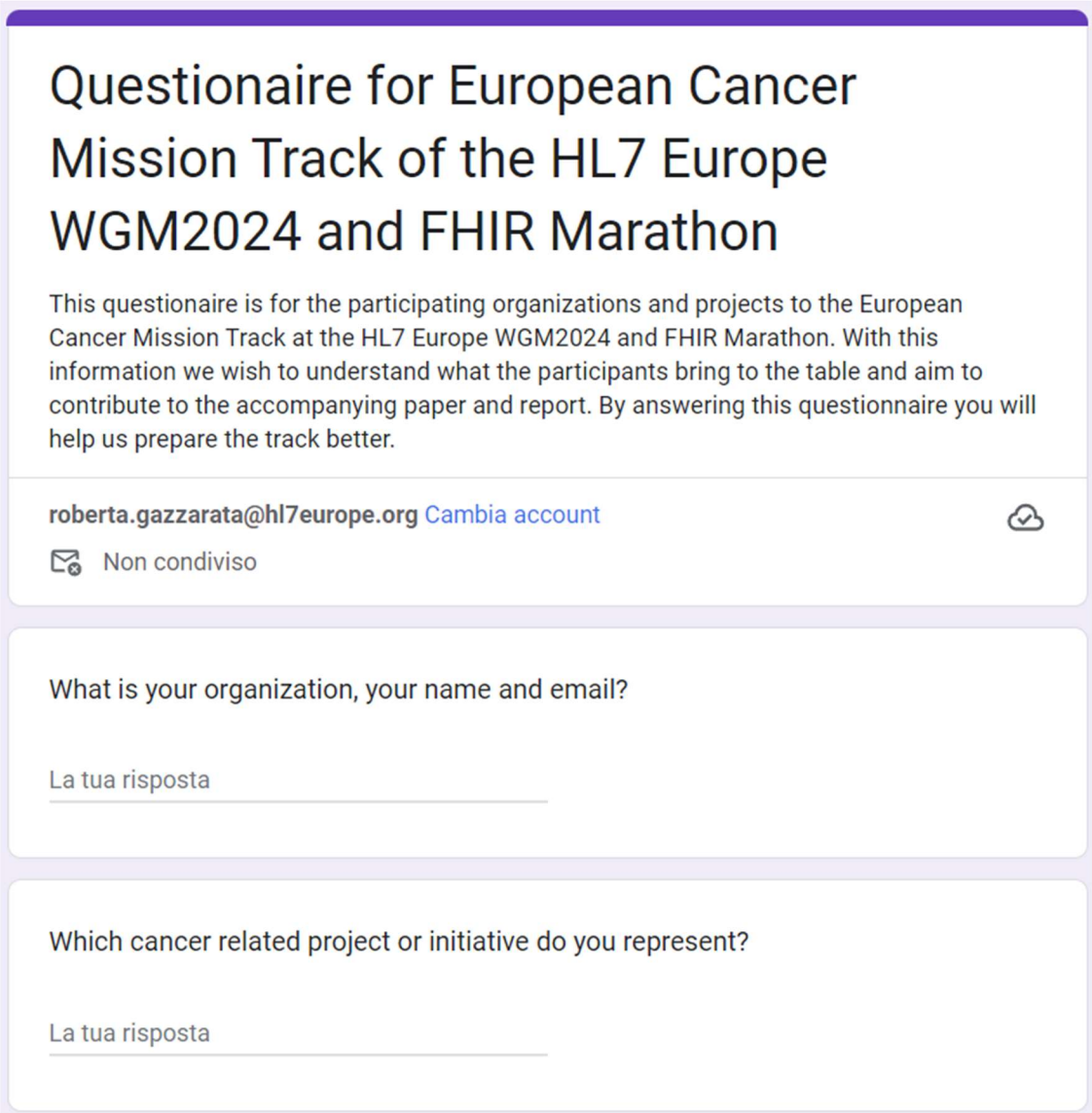


Figure 36 The questionnaire to prepare the "European Cancer Mission Track" at the HL7 Europe Working Group Meeting and HL7 FHIR Marathon in January 2024

The intent was to understand what the participants would bring to the table to help the organization of the track. In the questionnaire was asked:

- To provide:

[53] https://docs.google.com/forms/d/e/1FAIpQLSfqwFrJfxev3e6lSl2b0Mq1JkriRGSfiybi2LqnVWFmCxlJpw/viewform

    - Her/his name, mail and organization of the participant.
    - The name of the project or initiative in which the participant is involved and the type of cancer addressed by it.
    - The type of data with the participant works for the project or initiative (to choose between clinical data/patient summary, clinical data/laboratory results, clinical data/imaging reports, clinical data/prescriptions, clinical data/hospital discharge report, genomic data or to indicate others).
    - List reference relevant to the track.

- To indicate
    - If the project or initiative in which the participant is involved
        - Has developed a Common Data Model or an Ontology.
        - Has developed a FHIR Implementation Guide (IG) and, in a positive case, to provide the link to it.
    - If the participant works with OMOP and, in a positive case, if she/he is involved in OMOP to FHIR or FHIR to OMOP initiative.
    - If the participant addresses data quality and completeness.
    - How the participant expects to contribute to the track choosing between demonstrate FHIR IG, advance alignment among IGs, collaborate with other initiatives, work towards an aligned common data model for HL7 FHIR, create European version of mCODE.
    - What the main challenges the participant is facing in the project or initiative.

A HL7 Working Group Meeting and HL7 FHIR Marathon features hands-on FHIR development and testing and it is a chance to get hands dirty and learn by helping evolve the FHIR specification. Whether the format is virtual or in-person, implementers and developers can gain experience developing FHIR-based solutions and exchange data with other FHIR interfaces. Participants can engage in hands-on, heads down development and testing. There is an opportunity to work directly with other FHIR developers and senior members of the FHIR standards development team. Depending on the implementation state of the activities, participants can provide their experience in the specific track domain (at the starting phase, as happened for the “European Cancer Mission Track”) or bring some software intended to demonstrate FHIR connectivity.

`

An HL7 FHIR Marathon or Connectathon track lasts 3 days and the first day usually starts with the presentation of the proposed agenda and of the context in which the track is going to work. In the first day afternoon the track starts to work, and the activities end on the last day with the discussion of the results and the next steps, in the morning, and with the submission of the track report to HL7 Int, in the afternoon. For the "European Cancer Mission Track", the proposed agenda was the following:

- First day (16th January):
  - Morning: Presentation of the European/national projects and initiatives.
  - Afternoon: Definition of the scenarios which should work together and comparison between FHIR IGs developed in the different project.
- Second day (17th January):
  - Morning: Discussion of the first results and writing of the track report.
  - Afternoon: Definition of the scenarios which should work together and comparison between FHIR IGs developed in the different project.
- Third day (18th January):
  - Morning:Discussion of the first results, closing of the track report, definition of the next steps.
  - Afternoon: Submission of the track report.

The real agenda was a little different from the scheduled one. As proposed, on the first day morning IDEA4RC project, FLUTE project, OSIRIS, PanCareSurPass project and INCISIVE project were introduced. After the first feedbacks, in the afternoon, HL7 Europe decided to split the track in two sub-tracks: the first with the aim to present and introduce the structure of a FHIR IG and to compare IDEA4RC and PanCareSurPass FHIR IGs and the second with the goal to define a possible starting point for the creation of a first very preliminary draft of a HL7 Europe Cancer Common FHIR Implementation Guide.

To define a FHIR IG it is fundamental to have a data model on which base and then map the FHIR profiles created for the specific purpose. The OSIRIS data model was selected as the logical model to start, on the second day, the implementation of the FHIR IG draft. The paper

"A federated learning system for precision oncology in Europe: DigiONE"[54] was presented to the track as a publication that introduces the MEDOC (Figure 37).

Figure 37 The paper presented during the "European Cancer Mission Track" on MEDOC

Considering the relevance of the activity performed to create MEDOC, also highlighted by the journal on which it was published just few days before the HL7 Working Group Meeting and HL7 FHIR Marathon (i.e. Nature Medicine journal), HL7 Europe, in accord with the IDEA4RC coordinators, decided to select the data concepts defined in MEDOC of DIGIONE as a starting point for a comparison between the data adopted by the different European/national initiatives represented by the track participant.

[54] Mahon, Piers, et al. "A federated learning system for precision oncology in Europe: DigiONE." Nature Medicine (2024): 1-4

During the second day, a shared spreadsheet, "Comparison between Models v1"[55], was defined and the comparison was started. It was the base of a 2-phase surgery as it will be described. The spreadsheet was organized in two main sheets: "Clinical Data" and "Project Contacts". In the "Clinical Data" sheet (Figure 38) in the first 2 columns are reported the MEDOC categories and concepts as presented in Table 1. Then for each project or initiative 2 columns were prepare:

- "Present", to indicate if the corresponding MEDOC concept has been used in the project or initiative
- "Description", to provide some details on its adoption in the project or initiative or to indicate where it was adopted in data model or in FHIR resources elements.

These columns are useful to identify what elements should be mandatory in the data model and to understand what are the most adopted MEDOC concepts on which to start to work for the creation of the European Cancer Common Data Model and the corresponding European Cancer FHIR Implementation Guide.

MEDOC concept		OSIRIS		OMOP		IDEA4RC	
1. Demographics		Present	Description	Present	Description	Present	Description
	1.1 Date of birth (month)	YES	Present in Patient.BirthDate	YES	Present in Person.Birth_Datetime, Person.Year_of_birth, Person.Month_of_birth, Person.Day_of_birth	Yes	It is represented at CancerEpisode.AgeatDignosis
	1.2 Sex	YES	Present in Patien.Gender	YES	Present in Person.Gender_concept_id	Yes	Patient.sex
	1.3 Weight (with timestamp)	YES	Present in Analisys->Study->Series.Patient weight and Analisys.Date	YES	Present in Measureament	No	We only have Patient.BMI, due to anonymization
	1.4 Height	??		YES	Present in Measureament	No	We only have Patient.BMI, due to anonymization
	1.5 Healthcare ID (or other unique identifier)	NO	Name present in Present in Analisys->Study->Series.Institution name	NO	Name present in CARE_SITE.Hospital name	yes	Hospital.name
	1.6 Legal basis for data processing	YES	Consent	NO		No	Not present (for now at least)

Figure 38 The initial structure of the sheet "Clinical Data" of the "Comparison between Models v1"

In the "Project Contacts" sheet prepared during the "European Cancer Mission Track" in Athens, HL7 Europe, in accord with the IDEA4RC coordinators, prepared 5 columns: "Project", to indicate the name of the project or initiative, "FHIR IG", to provide the link to the defined FHIR IG in the project or initiative, "Affiliation", "Name" and "Mail" to collect the participant information.

[55] https://docs.google.com/spreadsheets/d/1LT6uDBI2ADlac5HuJ7Lhu1eillq6aHhT/edit?usp=sharing&ouid=101249583302891288660&rtpof=true&sd=true

Project	FHIR IG	Affiliation	Name	Mail
INCISIVE	https://simplifier.net/INCISIVE/~introduction, https://incisive-project.eu/wp-content/uploads/2023/08/INCISIVE_D3.4_StandardizationSuggestions_v1.0_FinalVersion.pdf	TIC Salut Social Foundation	Sara Alabart-Martínez, Shulei Huang	salabart@tics
IDEA4RC	https://build.fhir.org/ig/hl7-eu/idea4rc/	HL7 EU	Roberta Gazzarata	roberta.gazza
		UPM	Eugenio Gaeta	eugenio.gaet
		Istituto Tumori Milano	Annalisa Trama	annalisa.tram
		Istituto Tumori Milano	Paolo Lasalvia	paololasalvia
		Istituto Tumori Milano	Roberto Lillino	roberto.lillino
FLUTE	https://build.fhir.org/ig/hl7-eu/flute/	HL7 EU	Luc Chatty	luc.chatty@h
EDS Institut Paoli Calmettes	/	Fyrstain/ HOF		
Grand hopital de Charleroi data sharing	/	Fyrstain/ HOF		

Figure 39 The initial structure of the sheet " Project Contacts" of the “Comparison between Models v1”

3.2.2 From HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024 to the Medical Informatics Europe 2024

Taking into account the results obtained and presented in the chapter 4.2.1, after the first “European Cancer Mission Track” in Athens, HL7 Europe, in accord with the IDEA4RC coordinators, decided to continue the activities starting from the possible definition of a European Cancer Common Data Model and defined scope and principles of this WG. In detail its scope was to define a minimal, extensible, non-exhaustive European cancer data set, agnostic to the type of cancer, to be used across different use cases, leveraging on the experiences of the European projects working with primary and secondary usage and taking into account the availability and usability of reliable data in the EHR systems; and to specify an implementable HL7 FHIR representation of this data set. As principles, HL7 Europe and the IDEA4RC coordinators decided that the activity should be inclusive and cross-cutting, that means transversal to different cancer domains and purposes of use and to consider the needs of different communities; the approach should be incremental, starting with a minimal core set and then extending / enhancing the model in following iterations; the dataset should be designed for facilitating the representation with HL7 FHIR.

As a starting point, HL7 Europe decided to write a working paper to report the activities performed. The working paper was written by HL7 Europe in collaboration with all the participants of the “European Cancer Mission Track” at the HL7 Europe Working Group

Meeting and HL7 FHIR Marathon 2024. This working paper was very useful to collect information to report in chapter 2 of this document.

In addition, to exceed the "European Cancer Mission" WG to all the possible interesting people, HL7 Europe decided to propose a workshop at the 34th Medical Informatics Europe (MIE) Conference organized by the European Federation of Medical Informatics (EFMI).

3.2.3 Participation to Medical Informatics Europe (MIE) Conference 2024

MIE workshops are sessions where Medical Informatics Europe conference participants engage in discussion and activity concerning innovative topics. HL7 Europe and the IDEA4RC coordinators wrote a workshop proposal to submit to the MIE conference portal indicating, as required by MIE, an abstract, the introduction of the topic, the aim and rationale of the proposal, the expected outcome and a proposed programme in addition to a brief CV of the presenters. The proposal was accepted and in September 2024 the workshop "Towards a European Cancer Minimum Data Model and HL7 Europe Cancer Common FHIR Implementation Guide in the EHDS" was hosted by MIE conference at Athens.

Figure 40 The MIE conference logo

The agenda of the one-hour workshop was the following:

- Welcome and agenda
- The IDEA4RC project and its goals
- The OSIRIS initiative

- The IDEA4RC use cases and architecture
- The IDEA4RC Data Model and connection to OMOP
- The EEHRxF: HL7 FHIR IGs in progress and plans
- The HL7 Europe Working Group "European Cancer Mission": current status and plans
- Round of discussion and Q&A

3.2.4 From MIE 2024 to HL7 Europe Working Group Meeting and EU-a-thon 2025

After the positive experience at MIE Conference 2024, HL7 Europe, in accord with the IDEA4RC coordinators, decided to continue the activity on the comparison of the data models, by performing a second survey phase. In particular, also considering the experience derived by the activities of the other WPs (in particular WP2), the "Comparison between Models v1"[56] shared spreadsheet was changed by adding some columns and a new sheet.

In the sheet "Clinical Data" 2 columns for each project or initiative were added (Figure 41):

- "Code System", to indicate the vocabulary adopted to represent the specific MEDOC concept (if used in the project or initiative),
- "N centers out of XX with variable available", to indicate the number of centers with respect to the total number of centers involved in the project or initiative in which the data are available.

The aim of these columns was to try to understand if there are some common terminologies that can be adopted as reference one in the possible corresponding FHIR profiles and to check not only common relevant information across projects but also the real availability of this information at the hospital level. In fact, a European Cancer Common Data Model should be based not only on a wishful list of items considered relevant by experts but also balanced with the feasibility of collecting the data at the hospital.

56 https://docs.google.com/spreadsheets/d/1LT6uDBI2ADlac5HuJ7Lhu1eillq6aHhT/edit?usp=sharing&ouid=101249583302891288660&rtpof=true&sd=true

MEDOC Conept		IDEA4RC			
1. Demographics		Present	Description	Code System	N centres out of XX with variable available
	1.1 Date of birth (month)	Yes	It is represented at CancerEpisode.AgeatDiagnosis		
	1.2 Sex	Yes	Patient.sex	OMOP Gender	
	1.3 Weight (with timestamp)	No (BMI)	We only have Patient.BMI, due to anonymization	SNOMED	
	1.4 Height	No (BMI)	We only have Patient.BMI, due to anonymization	SNOMED	
	1.5 Healthcare ID (or other unique identifier)	yes	Hospital.name (ID)		
	1.6 Legal basis for data processing	No	Not present (for now at least)		

Figure 41 The new structure of the sheet "Clinical Data" of the “Comparison between Models v1”

In the sheet “ProjectContacts” 6 columns for each project or initiative were added (Figure 42):

- “Project objectives”, to summarize the project or initiative objectives
- “Primary or secondary use of data?”, to indicate if the project or initiative purpose is primary or secondary data use
- “Use cases for the project”, to summarize the use cases defined by the project or initiative
- “Is the data model proprietary?”, to indicate if the data is proprietary
- “Is the data model mapped to a standard one (FHIR, OMOP, openEHR etc)? Can you provide the data model and the map (if available)?”, to indicate if the data model was also mapped to standard data model/tools
- “Are there any redundancy in the model to facilitate the query for the clinicians?”, to indicate if it was necessary to insert some redundancy in the data model to facilitate the retrieval of information asked by clinicians.

The idea of adding these columns was moved to better understand the context of each project, the possible common use cases on which start to work on and to retrieve some information on the data model adopted and in the use of standards.

Project	Project objectives	Primary or secondary use of data?	Use cases for the project	Is the data model proprietary?	Is the data model mapped to a standard one (FHIR, OMOP, openEHR etc)? Can you provide the data model and the map (if available)?	Are there any redundancy in the model to facilitate the query for the clinicians?
MEDOC	- re-use of **secondary** data for natural history and outcome studies.	Secondary	Solid tumours, ongoing studies on breast, lung and colorectal	No	Mapped to OMOP	
IDEA4RC	- Secondary use of data for natural history, prognostic and predictive studies and studies on treatment effectiveness and quality of care.	Secondary	neutrophils/lymphocytes ratio (NLR) and prognostic index combining serological and inflammatory factors (PISIF) in primary retroperitoneal sarcomas - Identification of predictors of outcome after surgical treatment (with respect to both short term morbidity, survival, recurrences and quality of life) in H&N cancers. - Assessment of the outcomes (overall survival, disease free survival) of sino-nasal cancer patients treated with induction chemotherapy. - Induction chemotherapy is a treatment approach where chemotherapy is administered before the main treatment, such as surgery or radiation therapy." - Assessment of the role of photon and proton-based radiotherapy on the outcomes (overall survival, disease free survival) of low and intermediate grade mucoepidermoid cancers of	No	It's mapped to both FHIR and OMOP https://build.fhir.org/ig/hl7-eu/idea4rc/ OMOP mapping is in the DM file https://docs.google.com/spreadsheets/d/1ANErBpHQAW6ngn1kq-a7rPpeTosG-z2PHnwfUT6IUKI/edit?usp=sharing	Yes

Figure 42 The new columns of the sheet "Project Contacts" of the “Comparison between Models v1”

The sheet “New Variables” (Figure 43) was also added to allow each project or initiative to indicate potential variables not indicated in MEDOC, but that are relevant for the project or initiative. The considered columns were:

- “MEDOC Category”, to indicate the select the category of each new variable from the ones proposed by MEDOC
- “Project name”, to indicate the specific project or initiative
- “Additional variable useful for the project objectives and widely available in the hospitals”, to indicate each new variable
- “Description”, to provide some details on its adoption in the project or initiative or to indicate where it was adopted in data model or in FHIR resources elements.
- “Code System”, to indicate the vocabulary adopted to represent the specific MEDOC concept (if used in the project or initiative),
- “N centers out of XX with variable available”, to indicate the number of centers with respect to the total number of centers involved in the project or initiative in which the data are available.

The need of adding these sheets derives from feedback of WP2 which indicated that there were important variables for IDEA4RC that were not considered by MEDOC, so it could be possible that also other projects or initiatives had some ones.

MEDOC Category	Project name	Additional variable useful for the project objectives and widely available in the hospitals	Description	Code System	N centres out of XX with variable available
1. Demographics					
2. Clinical phenotype					
3. Biomarkers					
4. Treatment	ENCR	Cancer patients are not always treated in single hospitals. To understand care, referral patterns, we should aim to be able to reconstruct the patient journey			
5. Outcomes	IDEA4RC	- Disease status: progression, recurrence, stable disease, remission, complete remission - Disease extent: local, loco-regional, metastatic (already in) - Treatment response: related to disease status + margins after surgery	We want to track the status and the extent of the disease throughout its history. For that, having treatment responses is important, also, to understand decisions made by clinicians		

Figure 43 The new sheet "New Variables" of the "Comparison between Models v1"

In parallel, as will be discussed in chapter 3.2.5, HL7 Europe decided to organize a new cancer track called "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025" that was hosted in Lisbon on February 2025 to have the opportunity to have a second face to face event to continue the activities on the definition of an European Cancer Data Model. The intent was to arrive at the "Common Cancer Model Track" with a preliminary draft of a possible diagram of the entities and their relationship to be present, discuss and consolidate it during the official event.

For this reason, HL7 Europe, in accord with the IDEA4RC coordinators, decided to also ask the involved projects or initiatives to share a graphic representation of the data model adopted and or developed, if available and public. The reason of this request was that it could help to have a general overview of the data model to better understand the involved entities and their relationship to prepare the mentioned entity diagram.

Once the structure of the share spreadsheet "Comparison between Models v1"[57] was consolidated, HL7 Europe in December 2024 contacted all the involved projects or initiatives to continue to collaborate, by completing the spreadsheet, sharing the data model and

[57] https://docs.google.com/spreadsheets/d/1LT6uDBI2ADlac5HuJ7Lhu1eillq6aHhT/edit?usp=sharing&ouid=101249583302891288660&rtpof=true&sd=true

participating in the “Common Cancer Model Track” at the “HL7 Europe Working Group Meeting and EU-a-thon 2025”.

Taking into account the feedback collected during the Working Group Meeting of the 2 project cluster in which IDEA4RC is involved (i.e. InToEHR and EU-funded projects on data-driven approaches in cancer) to search for synergies with other projects in the clusters, the IDEA4RC coordinators decided to involve other projects. In detail, the IDEA4RC coordinators considered the projects that joined the European Commission Workshops “Landscaping data driven projects and initiatives in the cancer field – rationale and directions for better collaboration and integration”[58] and selected

- EHDS2Pilot
- EUCAIM
- AIDAVA
- canSERV
- EOSC4cancer
- I3LUNG
- ONCOVALUE
- BEACON
- DECIDER
- OPTIMA

In addition, they also decided to consider the PIONEER project. HL7 Europe contacted also these project contacts from the attender list of the workshops to ask to collaborate, by completing the spreadsheet, sharing the data model and participating in the “Common Cancer Model Track” at the “HL7 Europe Working Group Meeting and EU-a-thon 2025”.

To support the collaboration, HL7 Europe answered all the requests of help, also organizing different calls with single partners of some of these projects. HL7 Europe collected all the feedback and monitored the answer to collaborate, pressing contacts in case of a missing answer.

[58] https://digital-strategy.ec.europa.eu/en/library/report-workshops-landscaping-data-driven-projects-and-initiatives-cancer-field

3.2.5 Participation to HL7 Europe Working Group Meeting and EU-a-thon 2025

As announced in previous chapter, HL7 Europe decided to organize a new cancer track called "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025[59]" that was hosted in Lisbon in February 2025. In fact, considering the ongoing activities, HL7 Europe thought that it was the right time to have a face to face working session to start to design the European Cancer Data Model involving all the participants of the survey to join bringing their experience matured in their specific activities.

Figure 44 The "HL7 Europe Working Group Meeting and EU-a-thon 2025" poster

In detail, the scope of the "Common Cancer Model Track" was to make the first step to define a minimal, extensible, non-exhaustive European cancer data set, agnostic to the type of cancer, to be used across different use cases, leveraging on the experiences of the European projects working with primary and secondary usage and taking into account the availability and usability of reliable data in the EHR systems; the activity should be inclusive and cross-cutting, that means transversal to different cancer domains and purposes of use and to consider the needs of different communities; the approach should be incremental, starting with a minimal core set and then extending / enhancing the model in following iterations; the dataset should be designed for facilitating the representation with HL7 FHIR.

[59] https://hl7europe.org/hl7-working-group-meeting-2025/

All the results of the survey, especially the ones reported in chapters **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.**, were adopted to design, with the fundamental help of partners that worked on the IDEA4RC data model (WP2), a preliminary draft of a possible model of the entities and their relationship to be present, discuss and consolidate it during the official event. Considering that MEDOC is not a data model, HL7 Europe decided to select another starting point for the model design. Even if mCODE FHIR Implementation Guide cannot be used in Europe context and for secondary data use, as indicated in chapter 2.2, the great effort, guided by the American Society of Clinical Oncology, to define the entities and their relationships in the cancer management domains (Figure 45)[60] could be useful to propose a similar model applicable for the European context.

[60] https://build.fhir.org/ig/HL7/fhir-mCODE-ig/#overview

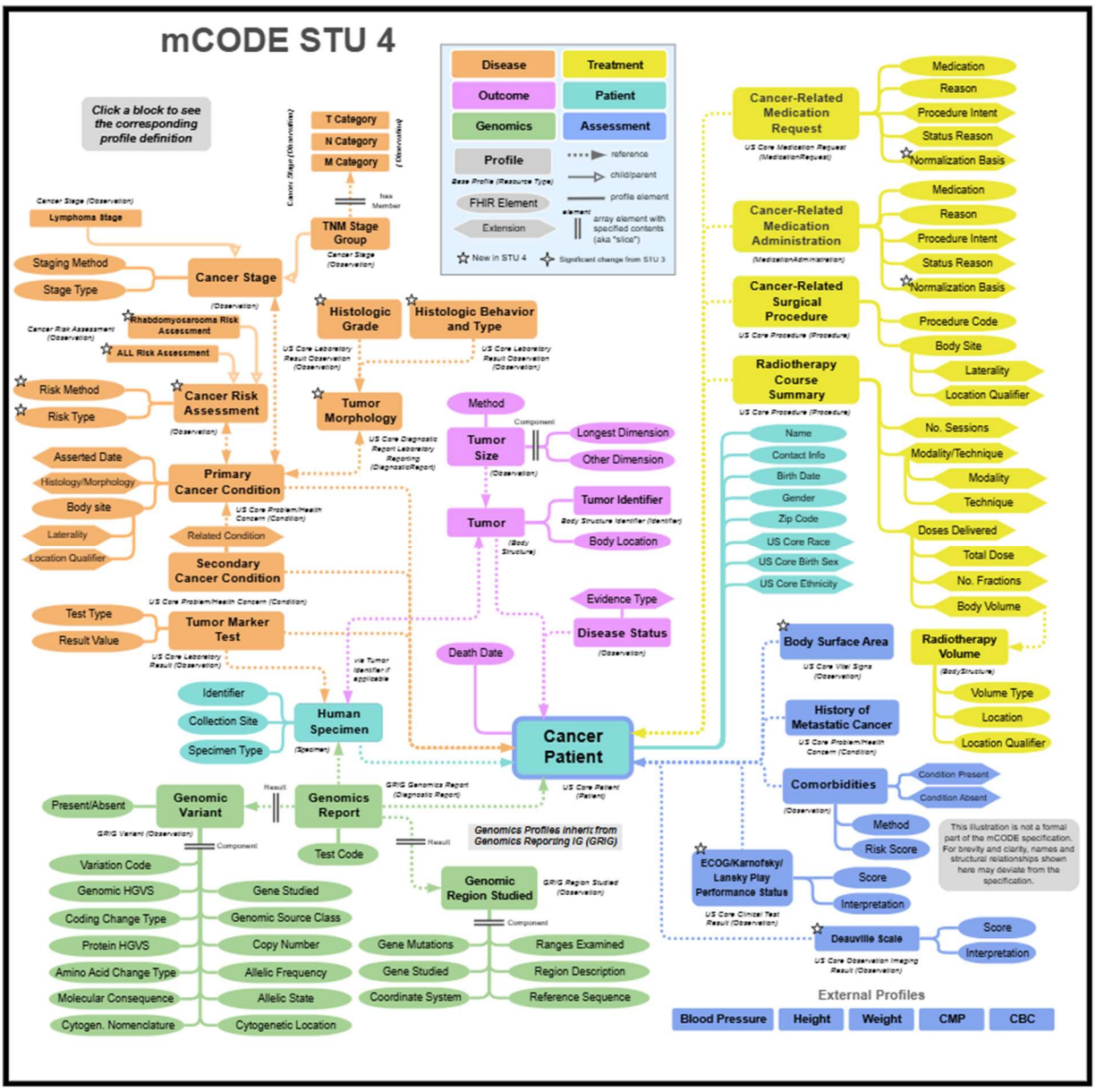


Figure 45 The model of entities and their relationship defined by mCODE

Therefore, starting from the model reported in Figure 45 by also considering input from other data models from the ones reported in Table 4, HL7 Europe prepared the draft of the model on which to start to discuss during the "Common Cancer Model Track". Considering the available time, the list of priority entities on which to start to work was obtained by the elaboration of the results obtained by the survey. A map between the MEDOC concepts and the corresponding entities in the prepared draft model and in mCODE model was prepared to guide the working section. In this activity, the contribution of the WP2 was fundamental. Several calls were organized to arrive at the final draft. In addition, during the discussions, HL7 Europe and WP2 partners agreed that to be able to reconstruct the patient journey and the history of the disease, a dedicated section should be planned to perform a similar exercise as the one done by mCODE to define a patient journey diagram (Figure 46) that should be reconstructed by the information that will form the European Cancer Common Data Model.

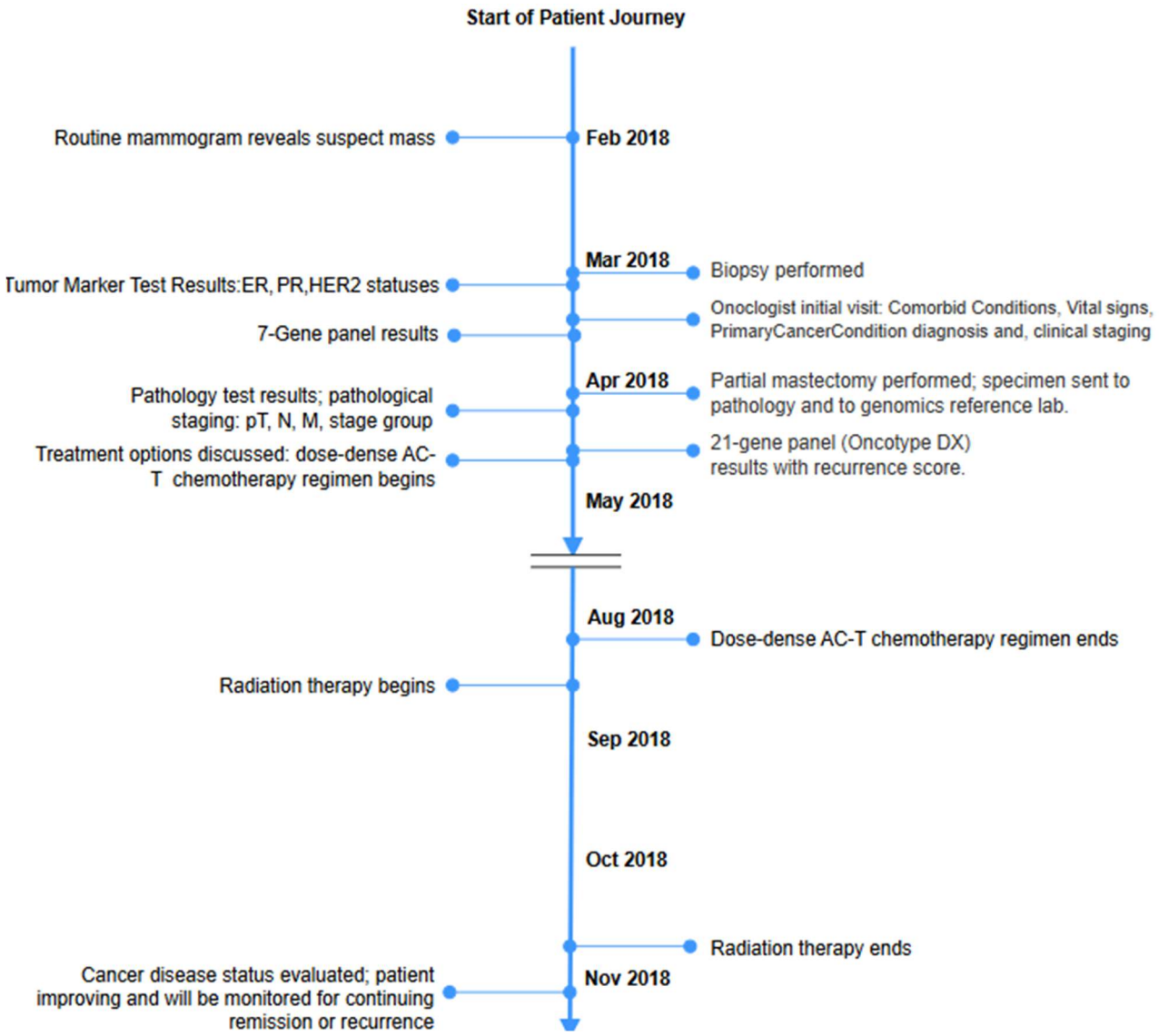


Figure 46 The example of a *patient*'s journey diagram defined by mCODE

Preparing the model draft, some crucial other points on which discuss during the "Common Cancer Model Track" emerged, as how to represent the metastatic cancer, disease status and the disease extent.

Taking into account all these aspects, the proposed agenda for the "Common Cancer Model Track" was the following:

- First day (12th February):
  - Welcome and attender registration
  - Agenda Presentation
  - Introduction:
    - Recap of the previous activities

- Introduction of the context: why we cannot use mCODE FHIR IG or OMOP?
- Presentation of the results of survey activities performed from the “European Cancer Mission Track” at the "HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024"

        - Aim and desiderata results
    - Working session 1: The cancer patient’s journey
        - Presentation of the journey
        - Discussion
    - Working session 2: The entity model
        - Presentation of the prepared material and the first draft
        - Discussion on the first draft and working in the draft defining entities and their relationship
        - Definition of the attributes of some of the defined entities
- Second day (13th February morning):
    - Recap of the results obtained in the previous sections
    - Discussion of the results
    - Conclusions and next steps

The proposed agenda was followed for the overall “Common Cancer Model Track”. On 12th morning the working session 1 was completed and the second one was started. The activities on working session 2 continued on 13th February morning. During the 12th morning the IDEA4RC data model was introduced and the methodology adopted to implement the IDEA4RC FHIR IG was presented as guideline to follow. In fact, as made for IDEA4RC, the correct approach before create a FHIR IG should be create and consolidate a data model able to satisfy the needs of the clinicians and the oncology researchers. The model could that be used from other SDO for implement other specifications.

Once the cancer patient’s journey (or better to call cancer journey as will be discussed in 4.2.5**Errore. L'origine riferimento non è stata trovata.**) defined in IDEA4RC was consolidated, the attention was focus on the creation of a timeline conceptual model starting from the prepared draft. For this activity the collaboration of the IDEA4RC partners attending the event

was fundamentals. For the design of the data model Excalidraw[61], a whiteboard tool to easily sketch diagrams that have a hand-drawn feel to them ,was used as done by WP2 in IDEA4RC.

At the end of 13th February morning HL7 Europe, in agree with the IDEA4RC coordinators, propose to the participants to continue the activities on the conceptual model by organizing monthly calls for March and April. The final aim was to finalize and consolidate the logical model to present and discuss it at international level (not only European) during a dedicated cancer track at the "HL7 International 2025 May Working Group Meeting & HL7 FHIR Connectathon" that will be hosted in Madrid in May 2025 and to ask to mCODE group of the CodeX FHIR Accellerator to collaborate to create an International Cancer Common Data Model.

[61] https://excalidraw.com/

4 RESULTS

In this section the results obtained are presented, starting from the ones got in the first period of activities based on the collaboration and participation to the HL7 Int initiatives Vulcan (FHIR to OMOP), and Codex (mCODE), and followed by the ones achieved in the activities performed at European level with the HL7 Europe Working Group “European Cancer Mission”.

4.1 CAMP FHIR presentation to IDEA4RC WPs, resolution of issues on terminologies adoption and collaboration in mCODE FHIR IG

The connection established between HL7 Europe and “OHDSI OMOP + FHIR” Working Group of the Vulcan FHIR Accelerator “FHIR to OMOP” project during the HL7 Int WGMs in January 2023 in Henderson (USA), allowed to collect artefacts and tools used, which then have been shared with the appropriate IDEA4RC WPs to collaborate in the design of IDEA4RC architecture. In particular, the CAMP FHIR platform and information about projects achieving mappings between Common Data Models were presented. The CAMP FHIR (Clinical Asset Mapping Program for FHIR)[62] is a software application developed by Green Team in collaboration with Orange Team as part of the Translator project, aimed at improving clinical data sharing within the Translator Consortium. The Biomedical Data Translator[63] (also known as "Translator") program was launched by the National Center for Advancing Translational Sciences (NCATS) to overcome challenges in application of the myriad biomedical datasets that are available today for the greater public good providing a comprehensive, open solution to accelerate clinical and translational research and drive innovations in clinical care and drug discovery. CAMP FHIR was motivated by the impossibility to define a unified agreed-upon clinical data model because different data models are present with specific purposes as OMOP or i2b2. To address these issues, CAMP FHIR employs HL7 FHIR as a "meta-CDM." or single standard to represent clinical data and efficiently transforms clinical data to FHIR irrespective of CDM, thus providing source-agnostic CDM-to-FHIR mapping. This involves two main steps (Figure 47):

1. mapping each source variable to its corresponding FHIR element
2. mapping each item in the source data’s value sets to the corresponding FHIR value set item, for variables with strict value sets

[62] https://researchsoftwareinstitute.github.io/data-translator/apps/camp-fhir
[63] https://researchsoftwareinstitute.github.io/data-translator/

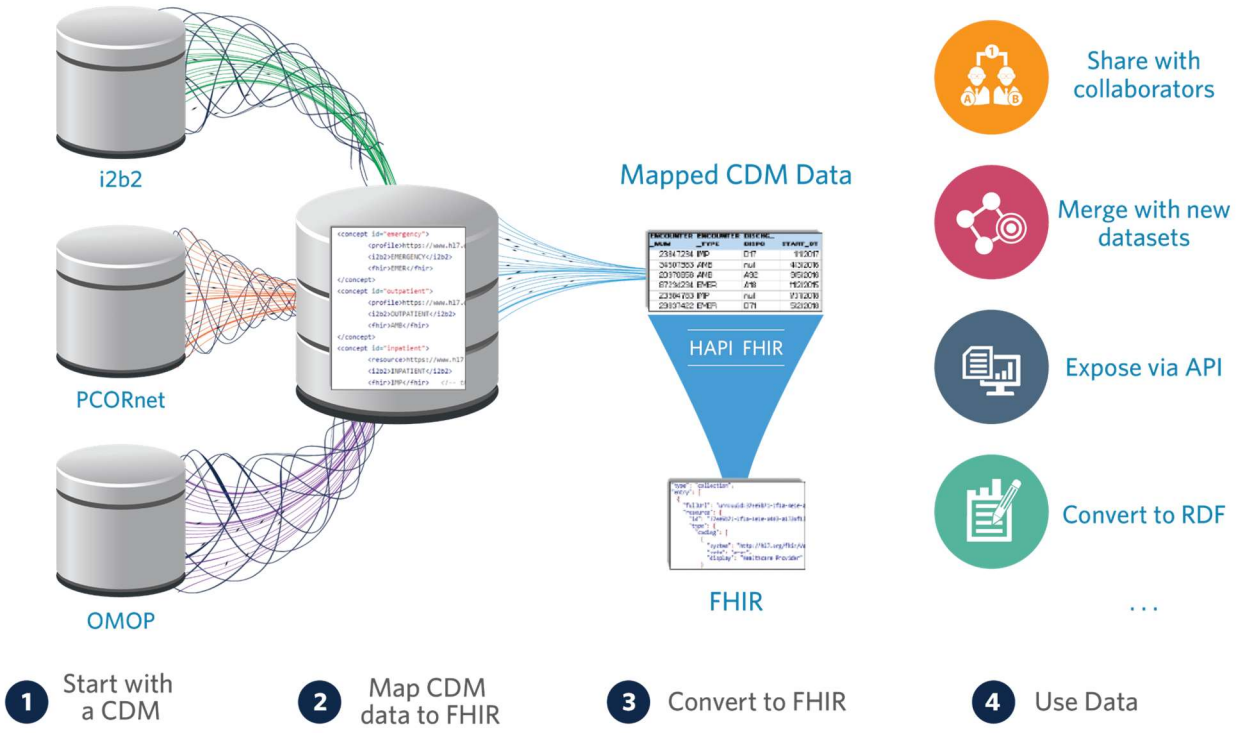


Figure 47 The CDM-to-FHIR mapping with CAMP FHIR

CAMP FHIR can be used to harmonize clinical data to save institutional resources over the alternative of standing up new CDMs to support multi-institutional collaborations. Moreover, using FHIR as a CDM could support rarer data sharing opportunities, such as collaborations between academic medical centers and community hospitals, thus democratizing participation and access.

During the "OHDSI OMOP + FHIR" WGM in January 2023 the CAMP FHIR was adopted to test and validate maps between FHIR-compliant Source Data and OMOP Relational Database as represented in Figure 48.

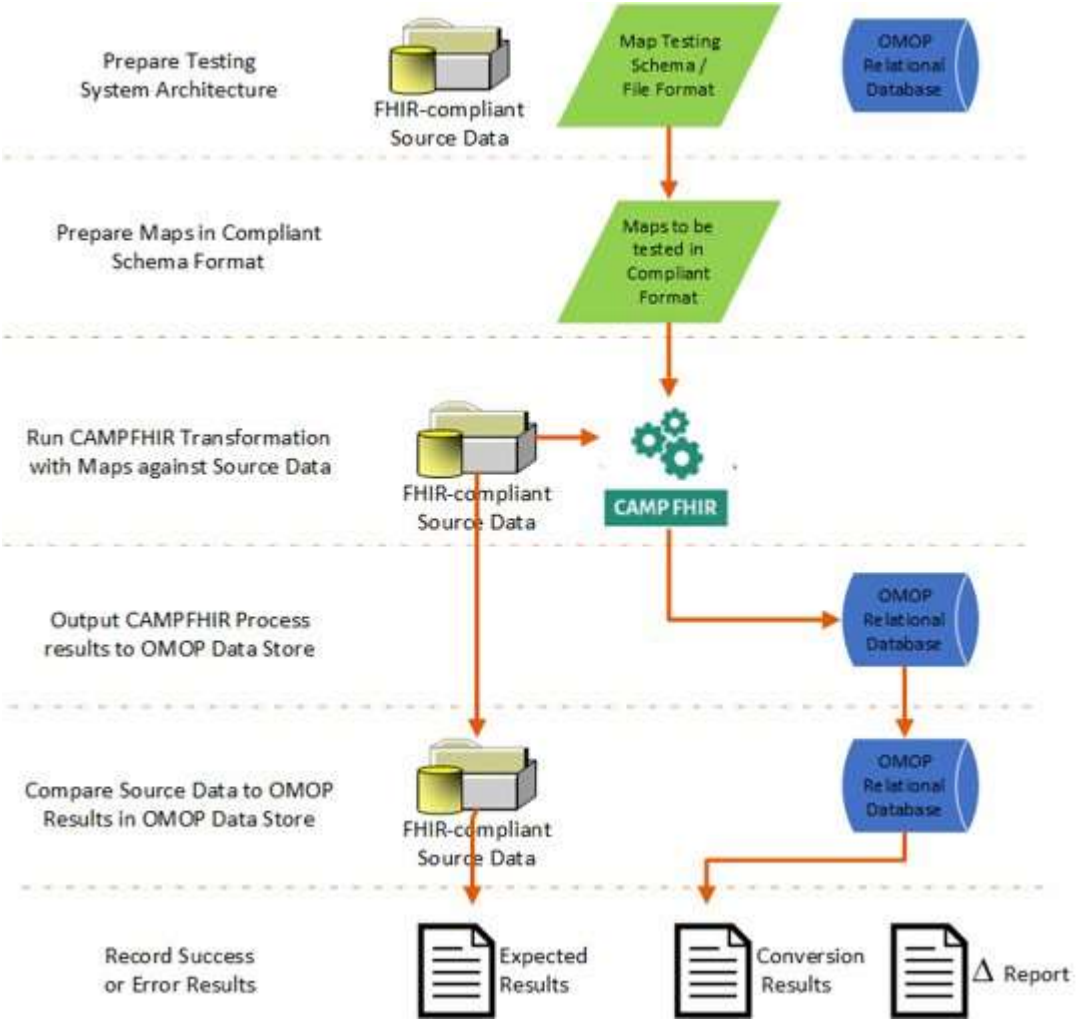


Figure 48 CAMP FHIR to test and validate maps between FHIR-compliant Source Data and OMOP Relational Database

After the meet during the HL7 Int WGMs in January 2023 in Henderson, the follow up connection between HL7 Europe and the WG allowed to discuss on some topics and issues based on the WP3 need of identifying OMOP artifacts (e.g. Athena concepts) in HL7 FHIR. It was necessary also to interact with other HL7 and OHDSI WGs especially on vocabulary topics were some of the open issues highlighted are under discussion (e.g. the HL7 TI WG meeting minute https://github.com/HL7/fhir/wiki/OMOP-terminology-discussion, or the OHDSI CDM Vocabulary Subgroup meetings where it has been identified that a) the FHIR community would like clarification on a versioning approach; b) needs a canonical URL for the OMOP Vocabulary that can be accessed through a REST API; c) requests additional review and input from members of the OHDSI vocabulary community).

During the conversations with the "OHDSI OMOP + FHIR" Working Group, OHDSI declared that there is a specific OHDSI WG that has among its committed objectives to produce a skeleton /

draft specification for OMOP / FHIR transformations based on the explorations and discoveries of its 4 Subgroups. and to develop a design for putting the OMOP Concept_ids on FHIR.

In this activity the HL7 community tool chat.fhir.org, described 3.1, was adopted in different situations. For example, it was used to explore the channel "OMOP + FHIR Terminologies" where there are specific conversations on the adoption of terminologies in OMOP and FHIR as the one "OHDSI/Athena FHIR system URL?" (Figure 49) that was joined by HL7 Europe.[64] An example of conversation that was created by HL7 Europe was "OMOP CDM 5.4 as FHIR LM" in the channel "OMOP + FHIR"[65]

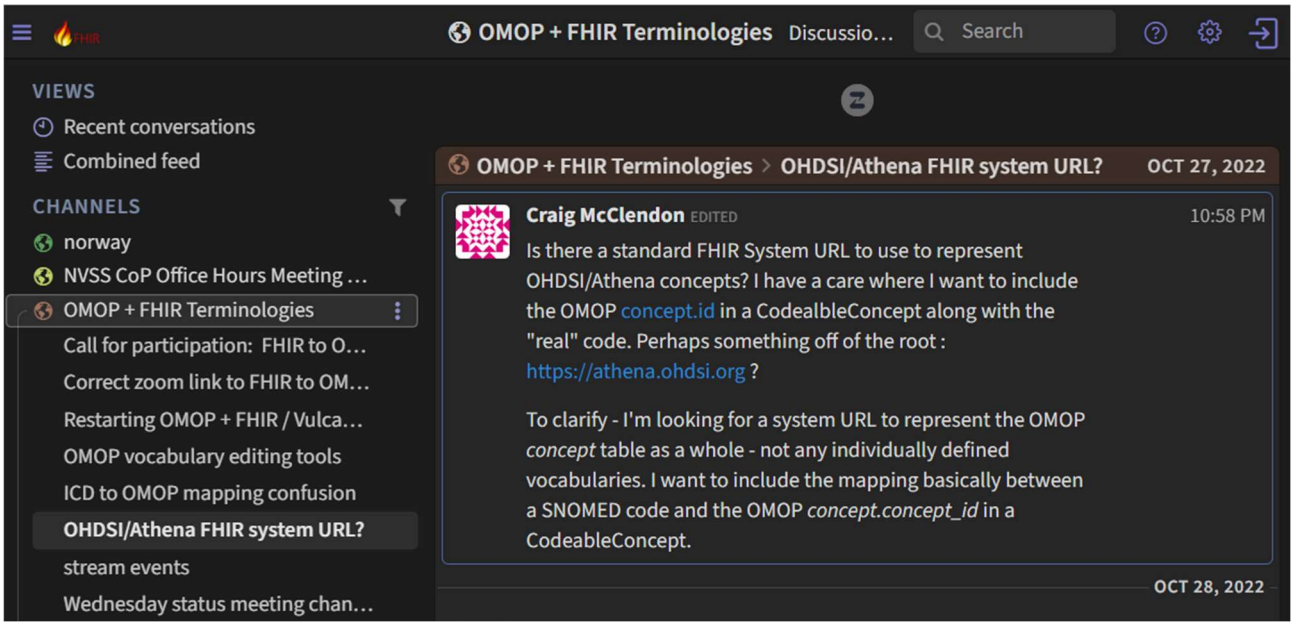


Figure 49 The conversation "OHDSI/Athena FHIR system URL? " in the chat.fhir.org channel "OMOP + FHIR Terminologies"

[64] https://chat.fhir.org/#narrow/stream/306130-OMOP-.2B-FHIR-Terminologies/topic/OHDSI.2FAthena.20FHIR.20system.20URL.3F
[65] https://chat.fhir.org/#narrow/stream/286658-omop-.2B-fhir/topic/OMOP.20CDM.205.2E4.20as.20FHIR.20LM

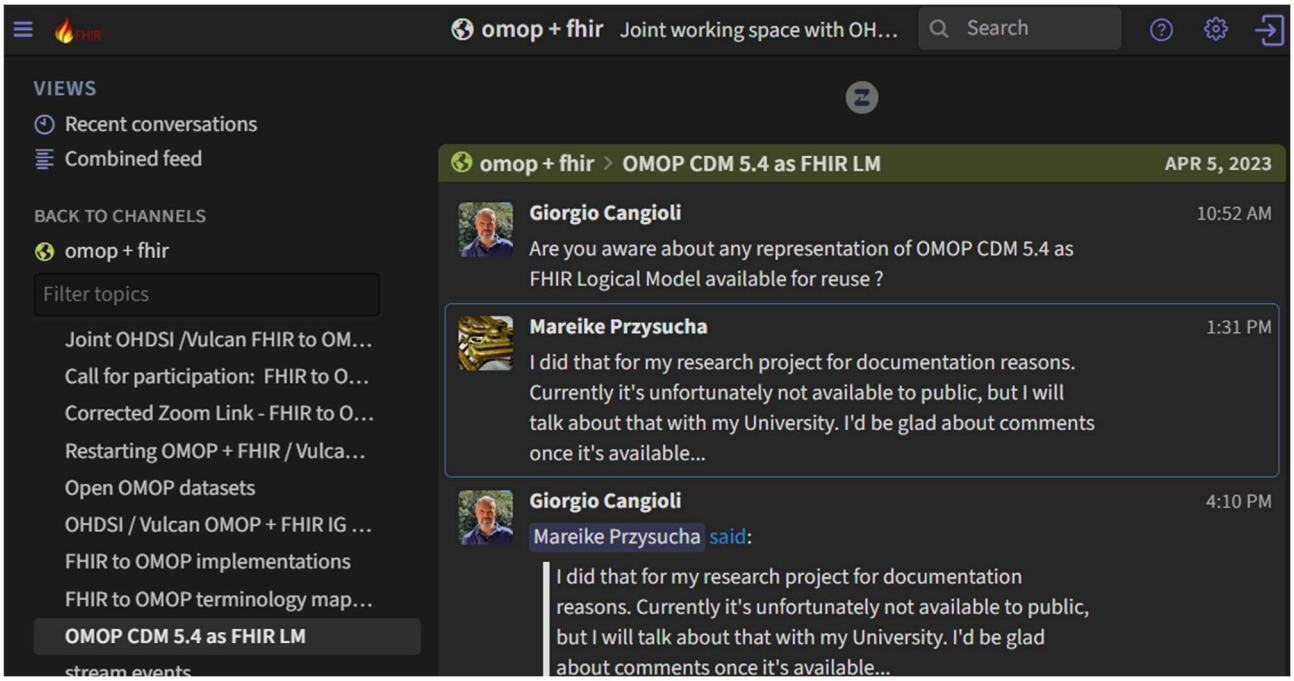


Figure 50 The conversation " OMOP CDM 5.4 as FHIR LM " in the chat.fhir.org channel "OMOP + FHIR"

As described in 3.1**Errore. L'origine riferimento non è stata trovata.**, another important collaboration during the activity period was with Codex FHIR Accelerator on mCODE FHIR Implementation Guide[66]. This IG was adopted only in some parts of the IDEA4RC FHIR Implementation Guide (see the Deliverable D3.1 "FHIR Implementation Guide" for more details) but in this adoption it was necessary to ask some changes to the mCODE FHIR IG based on the IDEAR4RC feedback. In particular, first a discussion "Question about radiotherapy summary profile"[67] in the channel "Cancer Interoperability" of chat.fhir.org was created to ask for some clarifications (Figure 51). From this discussion a formal Change Request "Deprecate TreatmentTerminationReason Extension"[68]for the mCODE FHIR IG (Figure 52) was triggered in HL7's Jira. The request was approved and became part of the Standard for Trial Use (STU) 3 Release of 26 October 2023 of mCODE FHIR IG[69]. Figure 53 shows the extract of the Release Notes of the STU 3 Release of 26 October 2023 related to the change request performed through HL7's Jira.

[66] https://hl7.org/fhir/us/mcode/index.html
[67] https://chat.fhir.org/#narrow/channel/179234-Cancer-Interoperability/topic/Question.20about.20radiotheraphy.20summary.20profile
[68] https://jira.hl7.org/browse/FHIR-41680
[69] https://hl7.org/fhir/us/mcode/STU3/change_log.html#use-statusreason-instead-of-treatmentterminationreason-extension-fhir-41680

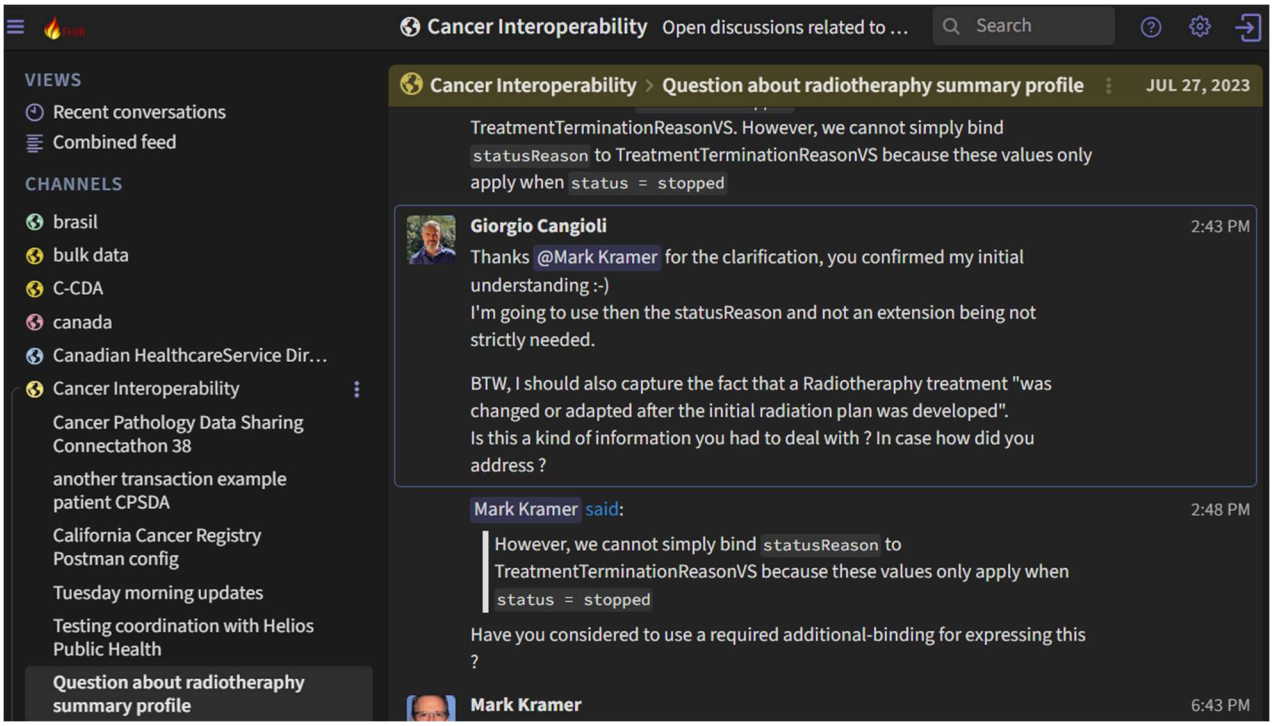
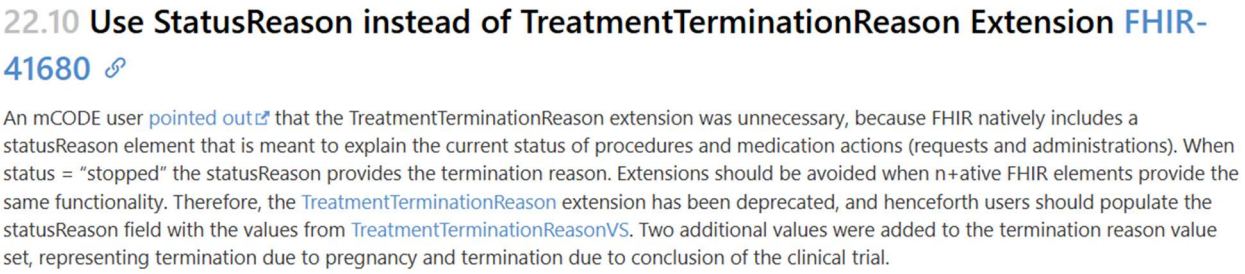


Figure 51 The conversation "Question about radiotheraphy summary profile" in the chat.fhir.org channel Cancer Interoperability

Figure 52 The Change Request "Deprecate TreatmentTerminationReason Extension" to the "US Minimal Common Oncology Data Elements (mCODE) (FHIR)" in HL7's Jira



22.10 Use StatusReason instead of TreatmentTerminationReason Extension FHIR-41680

An mCODE user pointed out that the TreatmentTerminationReason extension was unnecessary, because FHIR natively includes a statusReason element that is meant to explain the current status of procedures and medication actions (requests and administrations). When status = "stopped" the statusReason provides the termination reason. Extensions should be avoided when n+ative FHIR elements provide the same functionality. Therefore, the TreatmentTerminationReason extension has been deprecated, and henceforth users should populate the statusReason field with the values from TreatmentTerminationReasonVS. Two additional values were added to the termination reason value set, representing termination due to pregnancy and termination due to conclusion of the clinical trial.

Figure 53 The extract of the Release Notes of the STU3 release of the mCODE FHIR Implementation Guide

The collaboration with the FHIR community through of chat.fhir.org was also useful in the overall development of the IDEA4RC FHIR IG as reported in the conversations "Relapsed tumor in remission/resolved"[70] (channel "Cancer Interoperability") and "From site X to site Y"[71] (channel "implementers".

4.2 Towards the European Cancer Common Data Model

As mentioned in chapter 3.2**Errore. L'origine riferimento non è stata trovata.**, after the activities performed in the first period, in accord with the IDEA4RC coordinators, HL7 Europe decided to try to improve the primary and secondary data use in the context of the EHDS by creating a Working Group called "European Cancer Mission".

4.2.1 The questionnaire results, the Europe Cancer Common FHIR IG preliminary draft and the data model comparison survey (first phase) results

As previously indicated in chapter 3.2.1**Errore. L'origine riferimento non è stata trovata.**, to officially propose the "European Cancer Mission Track", some calls with possible participants were organized to create and complete a specific HL7's Confluence page[72]. In Figure 54 is reported a screenshot of the page where it is possible to see the proposal index.

[70] https://chat.fhir.org/#narrow/channel/179234-Cancer-Interoperability/topic/Relapsed.20tumor.20in.20remission.2Fresolved
[71] https://chat.fhir.org/#narrow/stream/179166-implementers/topic/From.20site.20X.20to.20site.20Y
[72] https://confluence.hl7.org/display/FHIR/2024+-+01+European+Cancer+Mission+Track

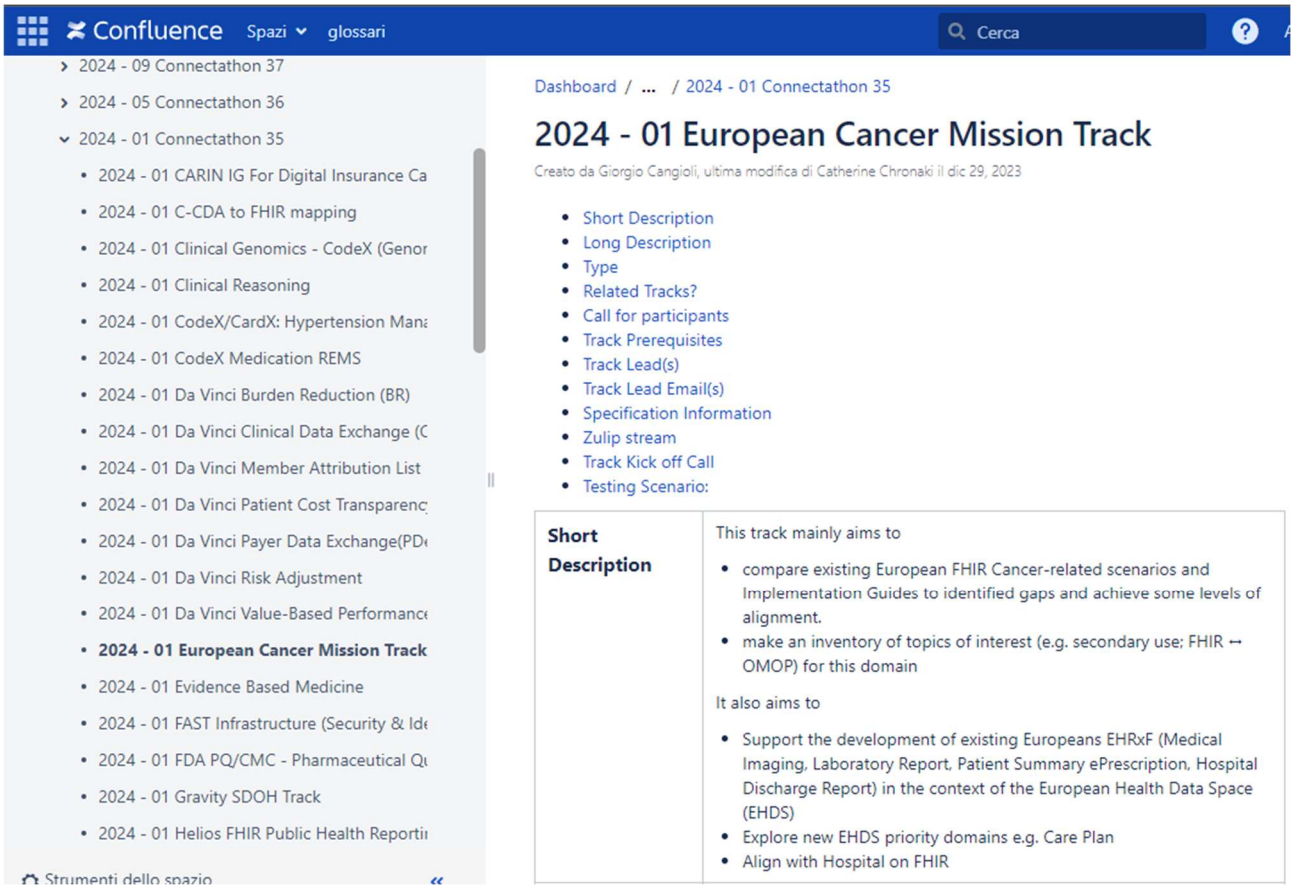


Figure 54 The "European Cancer Mission Track" proposal in the HL7's Confluence

After the approval of the proposed track, before the start of the "European Cancer Mission Track" a questionnaire[73] was prepared and proposed to the participating organizations and projects. HL7 Europe received at least an answer for all the projects that participated in the "European Cancer Mission Track" in Athens (except for FLUTE, but HL7 Europe is involved in this project, therefore an answer was not needed) for this project, indicating the great intent of all the participants. In detail, 11 participants to the "European Cancer Mission Track" completed the questionnaire before the event. A results HL7 Europe has obtained that:

- About 5 people were involved in IDEA4RC project (focused on the rare cancers head and neck, and sarcoma), 1 in PanCareSurPass project (focused on Childhood Cancer / Cancers diagnosed in pediatric patients), 1 in OSIRIS initiative (managing all cancer kind), 1 in INCISIVE project (focused on breast, lung, prostate and colorectal cancer), 1 in BBMRI-ERIC, 1 in IKNL Common Data Model for oncology (managing all cancer kind), 1 in no project.

[73] https://docs.google.com/forms/d/e/1FAIpQLSfqwFrJfxev3e6lSl2b0Mq1JkriRGSfiybi2LqnVWFmCxlJpw/viewform

- 9 people have worked with clinical data/patient summary, 6 with clinical data/laboratory results, 5 with clinical data/imaging reports, 6 with clinical data/prescriptions, 3 with clinical data/hospital discharge report, 4 with genomic data, 2 with radiomic/radiotherapy data, 1 with medical image data, 1 with cancer registry data, 1 with care plans, 1 with surgery and pathological report, 1 with classification of malignant tumor.
- 9 people have implemented a Common Data Model.
- 7 people have implemented a FHIR IG (for the list of FHIR IG implemented please see the next section).
- 7 people have worked or are going to work with OMOP and among them nobody was involved in OMOP to FHIR or FHIR to OMOP initiative.
- 7 people have addressed or are going to address data quality and completeness.
- 1 was expected to contribute to the track by demonstrating FHIR IG, 2 by advancing alignment among IGs, 8 by collaborating with other initiatives, 9 by working towards an aligned common data model for HL7 FHIR, 3 by creating a European version of mCode, 1 by bringing expertise from the provenance information domain, 1 by promoting OSIRIS CDM, 1 by promoting new standard code for cancer to be integrate into Athena vocabulary, 1 by presenting a video on the project and its application.
- Each participant is facing different challenges: lack of data maturity regarding stakeholders, extract structured data from free text, isolated views of participants in multi-organizational processes, consistency in the data model at many levels: mappings to FHIR and OMOP, clinical and technical coherence, the double mapping OMOP / FHIR for no standardized terminologies, balancing clinical needs with researcher needs, the different versions of FHIR and the use of national standards versus international, information not always available in OMOP (specific histology and topography combinations, specific sarcoma behaviours etc), limited structured information, the same data available in different data sources, limited experienced at the healthcare provider level, different types of data are coming from various hospitals systems and other kinds of systems.

To discuss these results:

- Most of the participants have mainly worked with clinical data/patient summary, clinical data/laboratory report, clinical data/prescription followed by clinical data/imaging report, genomic data, radiomic/radiotherapy data and finally only a

participant have experience in medical image data, cancer registry data, care plans, surgery and pathological report and classification of malignant tumor. This result is well aligned with the priority domains of the EEHRxF.

- Most of the participants, or the project in which they are involved, have implemented a Common Data Model highlighting the importance of defining a data model in each initiative. More than half of participants declared that in the project a FHIR IG was implemented, proving that the need of FHIR IG to support interoperability is clear.
- The same number of people indicated that they have addressed or are going to address data quality and completeness and they are working or are going to work with OMOP but nobody was involved in OMOP to FHIR or FHIR to OMOP initiative, indicating that the interest in the OMOP adoption is present but the map with FHIR could be improved.
- Considering the expectation in their contribution, the most selected answers were working towards an aligned common data model for HL7 FHIR, collaborating with other initiatives, followed by creating a European version of mCode, proving that the intent of the track to collaborate towards a common approach to the use of HL7 FHIR in cancer domain in Europe was clear to all the participants.

The "European Cancer Mission Track" in Athens, led by HL7 Europe, lasted 3 days in which the group hardly worked to take the first steps towards a European Cancer Common Data Model and a HL7 Europe Cancer Common FHIR Implementation Guide (Figure 55). The program was changed from the scheduled agenda. This is a common aspect for the first meeting of a new WG during the HL7 Working Group Meeting and HL7 FHIR Marathon. Before the start of the track, HL7 Europe did not think to split it in sub-track, but after the first day morning in after the presentation of the projects, HL7 Europe realized that an presentation of the scope and the structure of a FHIR IG was needed to be sure that all the participant have the base knowledge to start the activities and understand the intent of the WG. For this reason, HL7 Europe decided to divide the WG in 2 groups: one to perform this training and then to start to compare 2 FHIR IGs and another to work on the data model to define a possible starting point for the creation of a first very preliminary draft of a European Cancer FHIR IG. Considering the idea of adopting MEDOC to start a comparison between the data managed by each involved, HL7 Europe decided to continue the WG meeting in 2 parallel sections, one to work on the comparison activity and the other to implement a draft of FHIR IG.

Figure 55 The WG in action during the “European Cancer Mission Track”

As results, HL7 Europe obtained, on one hand, the preliminary draft of the HL7 Europe Cancer Common FHIR Implementation Guide defining Patient Identity and the primary cancer Condition, and, on the other hand, the shared spreadsheet mentioned in the chapter 3.2.1**Errore. L'origine riferimento non è stata trovata.**, “Comparison between Models v1.xlxs”[74] (please consider only columns “Present” and “Description”).

Referring to the shared spreadsheet, Table 2 represents in a schematic view the results obtained with this survey. Considering that it contains the overall results achieved also with the second phase of the survey, as described in chapter 3.2.4**Errore. L'origine riferimento non è stata trovata.**, the results of the “European Cancer Mission Track” have been reported in light blue to distinguish them from the second phase ones (represented in purple). During the event also a proposal of maps in OMOP was indicated, as reported in Table 2. Table 5 presents in light

74 https://docs.google.com/spreadsheets/d/1LT6uDBI2ADlac5HuJ7Lhu1eillq6aHhT/edit?usp=sharing&ouid=101249583302891288660&rtpof=true&sd=true

blue the list of FHIR Implementation Guides implemented by each project or initiative that joined the "European Cancer Mission Track".

HL7 Europe has found that there were elements present in all the involved projects as patient sex, primary cancer diagnosis and comorbidities, date of primary cancer diagnosis, treatment drug dose and start and end date, radiotherapy dose and start and end date, surgery type and date, participation in clinical trial. Some elements were present in most of the initiatives as methods of primary cancer diagnosis, disease stage in a recognized standard such as TNM, radiotherapy type, metastasis presence/absence and Vital status. These results are useful to identify what elements should be mandatory in the data model and to understand what are the most adopted MEDOC concept on which start to work for the creation of the European Cancer Common Data Model and the corresponding European Cancer FHIR Implementation Guide

During this first analysis some issues emerged. First of all, MEDOC is not a data model but a list of data concepts and some MEDOC concepts are single simple data (e.g. data of birth and sex), which are easy to understand and to map in a model, while other ones are more complex data (e.g. Primary cancer diagnosis and comorbidities), which require a more detailed analysis.

In addition, MEDOC was created to support the automation of outcomes research studies, and it is based on primary use of data (e.g. weight and height) that are reasonable EHR availability across Europe. However, usually the projects are based on secondary use of data and for privacy reasons they adopt pseudo-anonymised/anonymized data (e.g. Body Mass Index (BMI) instead of weight and height).

4.2.2 The "European Cancer Mission Projects: Towards a common approach to the use of HL7 FHIR" working paper

Taking into account the results obtained with the "European Cancer Mission Track" at the HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024, as indicated in chapter 3.2.2**Errore. L'origine riferimento non è stata trovata.**, HL7 Europe wrote a working paper. It was titled "European Cancer Mission Projects: Towards a common approach to the use of HL7 FHIR"[75] and was organized in the following sessions:

[75]https://docs.google.com/document/d/1xvNfG5SeqTJf7k_FplP3L5FW0AW509Tb-evfFiKQ-tw/edit?tab=t.0#heading=h.gh887qbgttv0

- Contributors: the list of the contributors with contribution type
- Abstract: paper summary
- Introduction: presentation of the context
- Background: brief description of the European/national projects and initiatives presented or relevant for the context
- Methodology: description of the methodology adopted
- Results: presentation of the achievements
- Discussion: discussion of the results
- Next Steps – Recommendations: presentation of the future activities and advise.

This shared document was also important for this deliverable because it was the starting point for its writing and for the proposal of the workshop at the MIE conference.

European Cancer Mission Projects: Towards a common approach to the use of HL7 FHIR

Working Paper: to be published in collaboration with EFMI

Contributors

List of people which are participating to the activity of the working group indicating:

- if they organized or participated to the "European Cancer Mission Track" at the HL7 Europe Working Group Meeting and HL7 FHIR Marathon in Athens on 15th - 19th January 2024 (WGMFM-ATH)
- if they create or complete the proposed questionnaire (Q)
- if they are contributing to write this Working Paper or the activity after the "European Cancer Mission Track" (WP)

Name	Affiliation	Email	Note
Catherine Chronaki	HL7 Europe	chronaki@hl7europe.org	WGMFM-ATH, Q, WP
Roberta Gazzarata	HL7 Europe, Healthtropy Srl	roberta.gazzarata@hl7europe.org	WGMFM-ATH, Q, WP
Giorgio Cangioli	HL7 Europe	giorgio.cangioli@gmail.com	WGMFM-ATH, WP

Figure 56 The working paper "European Cancer Mission Projects: Towards a common approach to the use of HL7 FHIR "

4.2.3 The workshop "Towards a European Cancer Minimum Data Model and European Oncology FHIR Implementation Guide in the EHDS" at MIE Conference 2024

The workshop "Towards a European Cancer Minimum Data Model and European Oncology FHIR Implementation Guide in the EHDS" was proposed by HL7 Europe, approved by MIE committee and was hosted by the MIE conference at Athens. About 30 participants to the MIE conference joined the workshop, which was a great opportunity to present IDEA4RC and the Working Group "European Cancer Mission". The EFMI community welcomed this initiative with enthusiasm, proving the interest of the activity and motivating it to continue.

4.2.4 The data model comparison survey (second phase) results

As previously indicated in chapter 3.2.4**Errore. L'origine riferimento non è stata trovata.**, after the participation to the MIE conference, a second survey phase was performed. The shared file "Comparison between Models v1"[76] was little changed to collect other information and the selected European projects were contacted to collaborate in the survey. Among the 11 selected and contacted European projects, HL7 Europe get an answer by 31 January 2025 from 8 ones (EHDS2Pilot through BBMRI-ERIC as pilot, EUCAIM, AIDAVA, EOSC4cancer, I3LUNG, ONCOVALUE, BEACON, OPTIMA); among these 8, 4 effectively collaborated (EHDS2Pilot, EUCAIM, AIDAVA, EOSC4cancer). The other 4 indicated that in this period they neighed have time nor intend to participate nor are not sure to be authorized to participate. Anyway, in general, they are interested in the topic and want to be updated on the activities that will be performed and the results that will be achieved. However, thanks to the communication activities (as the participation to MIE conference) and direct contact of IDEA4RC coordinators, the European Network of Cancer registry (ENCR), PROCANCER-I, 1+MG, ONCOFAIR projects (sometimes partially) collaborated in the survey.

Starting from the information collected in the sheet "Clinical Data" of the "Comparison between Models v1", in addition to the project that have already made it, 6 projects completed it in this second survey phase and HL7 Europe obtained another proposal of map in OMOP. Table 2 reports the overall results achieved with the survey: the ones derived from the first phase are represented in light blue, while the ones derived from this second phase are represented in purple. In addition to the results already discussed in chapter 4.2.1**Errore. L'origine riferimento non è stata trovata.**, the survey highlighted that the data models adopted in the different involved projects or initiatives are very heterogeneous. In fact, as Table 2 reports in column "Present in Project (with details)" the MEDOC concepts are mapped differently by each project or initiative. This was also confirmed by a very preliminary analysis performed on graphical representation of the data model implemented by the different projects that shared it with HL7 Europe (reported in Table 4). This is understandable because each project or initiative has put its effort into defining a data model that can satisfy the specific needs and user cases to achieve its goals (the objectives and the use cases are reported in Table 3). This confirms that an

[76] https://docs.google.com/spreadsheets/d/1LT6uDBI2ADlac5HuJ7Lhu1eillq6aHhT/edit?usp=sharing&ouid=101249583302891288660&rtpof=true&sd=true

additional effort, based on the experience of each project is needed to create a European Cancer Common Data Model that could help to facilitate the data sharing among the projects.

In addition, referring to sheet “Clinical Data” the HL7 Europe obtained that there are some MEDOC concepts that can be mapped in different ways in OMOP (e.g. Data of birth, Sex). Referring to data semantics, HL7 Europe founded that the are some MEDOC concepts (e.g. Sex, Primary Cancer diagnosis, Biomarker name) that are already coded with international vocabularies as SNOMED, LOINC, ICD-03 and it will be very useful to be selected as reference one for the European Cancer Common Data Model. For some concepts such as the Primary Cancer diagnosis that represents complex data, it is possible that different code systems should be adopted to represent the needed semantics. However, an additional deeper analysis could be performed in future, after that the entities that will compose the European Cancer Common Data Model will be defined.

The survey highlighted the difficulty to come to know if the data are available in the care facilities, in fact, only 2 projects indicated the MEDOC concepts for which the corresponding data are available (in any format). These aspects should be considered in the definition of the cardinality of the entities and of the corresponding attributes which will form the European Cancer Common Data Model.

`

MEDOC Categories	MEDOC concepts	Present in Project (with details)	Possible map in OMOP	Possible Code System
1. Demographics	1.1 Date of birth (month)	**OSIRIS** (clinical module Table Patient.BirthDate) **IDEA4RC** (CancerEpisode.AgeatDiagnosis) **PanCareSurPass** (Patient.birthDate (required)) **AIDAVA** (Patient.hasBirthDateTime) **EOSC4Cancer** (Person's year of birth) **EUCAIM** (Patient.BirthYear. (The month not)) **ECNR** **ONCOFAIR** (Patient.birthdate)	Observation.Date of birth Person.year_of_birth /Person.month_of_birth	ISO8601 (**OSIRIS**) SNOMED (**OMOP**)
	1.2 Sex	**OSIRIS** (clinical module Table Patient.Gender) **IDEA4RC** (Patient.sex) **BBMRI-ERIC** (Patient Data: SEX) **INCISIVE** (Gender: Patient.gender) **PanCareSurPass** (Patient.gender (required)) **AIDAVA** (Patient.hasAdministrativeGender) **EOSC4Cancer** (Person's sex at birth) **EUCAIM** (Patient.BirthSex) **ECNR** **ProCAncer-I** (Present in Person.gender_concept_id) **ONCOFAIR** (Patient.gender)	Observation.Gender Person.gender_concept_id	FHIR AdministrativeGender (**OSIRIS, IDEA4RC, INCISIVE, PanCareSurPass, ONCOFAIR**) SNOMED (**OMOP, AIDAVA, EOSC4cancer, EUCAIM**) OMOP Gender (**IDER4RC, ProCAncer-I**) LOINC (**EUCAIM**)
	1.3 Weight (with timestamp)	**OSIRIS** (Analisys->Study->Series.Patient weight and Analisys.Date) **AIDAVA** (Measurement.hasQuantity)	Measurement.Body Weight	SNOMED (**OMOP, IDEA4RC**)
	1.4 Height	**OSIRIS** **AIDAVA** (Measurement.hasQuantity)	Measurement.Body Height	SNOMED (**OMOP, IDEA4RC**)

	1.5 Healthcare ID (or other unique identifier)	**OSIRIS** (Analisys->Study->Series.Institution name) **IDEA4RC** (Hospital.name) **BBMRI-ERIC** (Patient Data: PATIENT_ID) **INCISIVE** (Patient Number: Patient.identifier and DataProvider: Organization.identifier) **PanCareSurPass** (Patient.generalPractitioner:primaryTreatmentCenter.identifier (optional)) **AIDAVA** (DataProvider.hasCode) **EOSC4Cancer** (Person identifier) **EUCAIM** (Patient.CareProvider and Patient.ManagingOrganisation) **ProCAncer-I** (Care_site.care_site_id and Care_site.care_site_name) **ONCOFAIR** (Patient.identifier)	Name present in Care_site.Hospital name ID in care_site.care_site_id and Name in Care_site.Hospital name	FINESS (French codification for healthcare ID) (**OSIRIS**) LOINC (**OMOP**)
	1.6 Legal basis for data processing	**OSIRIS** (Consent) **INCISIVE** (Present in the user interface) **AIDAVA** (Consent.hasStatusCode) **EUCAIM** (At a dataset level - not patient level (through DCAT-AP))	Observation as Clinical Finding: Consent status for record sharing - Concept ID 4185316	SNOMED (**OMOP**)
2. Clinical phenotype	2.1 Primary cancer diagnosis and comorbidities, typically in International Classification of Disease standards	**OSIRIS** (Primary cancer diagnosis present in Tumor pathology event.Morphology Code and Topography Code, comorbidities present in Related pathology.Pathology Code. There is not a Primary cancer element, all cancer diagnosis are managed with Tumor pathology event.) **IDEA4RC** (There is an overarching CancerEpisode to which are associated several Cancer Episode	Primary cancer diagnosis is present in Condition (Morphology/Topography). Comorbities (not specified as comorbidities, only as	ICD-O3 (**OSIRIS, PanCareSurPass, EUCAIM, ENCR**) SNOMED (**OMOP, IDEA4RC, INCISIVE, AIDAVA, EOSC4cancer, EUCAIM, ProCAncer-I**)

	such as ICD10, ICD9 or ICD-O-3	coded in agreement with Athena vocabulary whenever possible) **BBMRI-ERIC** (Present in Histopathology: HIST_LOCALIZATION) **INCISIVE** (Presento nly for patient with cancer) **PanCareSurPass** (Present in Condition.code.coding:iccc3-classification (required) and Condition.histologyMorphologyBehavior.value (required)) **AIDAVA** (ProblemCondition.hasCode) **EOSC4Cancer** (CRC Diagnosis) **EUCAIM** (PrimaryCancerCondition.code, PrimaryCancerCondition.BodySite, Comorbidities.code.) **ECNR** **ProCAncer-I** (Condition_occurrence.condition_concept_id)	condition) are present in Condition Primary cancer diagnosis is present also Episode as Episode of Care.	ICCC-3 (**PanCareSurPass**) ICD-10 (**AIDAVA, EUCAIM**)
	2.2 Charlson comorbidity index (derived from 17 comorbidities in 2.1)	**IDEA4RC** (Patient.Charlson Comorbidity index)	Measurement.Charlson comorbidity index	SNOMED (**OMOP, IDEA4RC**)
	2.3 Date of primary cancer diagnosis	**OSIRIS** (Tumor pathology event.Diagnosis Date) **IDEA4RC** (CancerEpisode.DateofDiagnosis), **BBMRI-ERIC** (Patient Data: DATE_DIAGNOSIS) **INCISIVE** (Age at diagnosis: DiagnosticReport.extension)	Episode.episode_start_date Condition.condition_start_date	ISO8601 (**OSIRIS**) SNOMED (**OMOP, EUCAIM**)

	PanCareSurPass (Condition.extension:assertedDate (optional)) **AIDAVA** (ProblemCondition.hasRecordDateTime) **EOSC4Cancer** (Mesurement date) **EUCAIM** (PrimaryCancerCondition.dateOfDiagnosis) **ECNR**		
2.4 Method of primary cancer diagnosis	**OSIRIS** (Tumor pathology event.Analisys.Type) **IDEA4RC** (CancerEpisode but only for head and neck and sarcoma), **BBMRI-ERIC** (Diagnostic exam) **INCISIVE** (Procedure resource) **EUCAIM** (Procedure.Category) **ECNR** **ProCAncer-I** (link to procedure_occurrence)	Present in Procedure_Occurrence	SNOMED (**OMOP, IDEA4RC, INCISIVE, EUCAIM, ProCAncer-I**)
2.5 Performance status (for example, coded by ECOG or Karnofsky standards)	**OSIRIS** **IDEA4RC** (Patient.ecogPS, Patient.KarnofsyIndexAtDiagnosis) **BBMRI-ERIC** (Diagnostic exam) **INCISIVE** (Preformance mesure status(ECOG): Observation.component.valueInteger) **EUCAIM** (ECOGPerformanceStatus.code)	Observation or Measurement	SNOMED (**OMOP, IDEA4RC, IDEA4RC, EUCAIM**)
2.6 Disease stage in a recognized standard such as TNM	**OSIRIS** (Tumor pathology event.TNM) **IDEA4RC** (Stage (Different for head and neck and sarcoma)) **BBMRI-ERIC** (Histopathology - TNM)	Observation.Pathologic TNM stage Measurement	TNM (**OSIRIS**) SNOMED (**OMOP, IDEA4RC, EOSC4cancer, EUCAIM**)

		INCISIVE (Preformance mesure status(ECOG): Observation.component.valueInteger) **AIDAVA** (TNMClassification.hasValue) **EOSC4Cancer** (CRC pTNM) **EUCAIM** (CancerStage.code, CancerStage.method, CancerStage.value) **ECNR** **ProCAncer-I** (Measurement.measurement_concept_id)		OMOP Cancer Modifier (**OMOP, IDEA4RC, ProCAncer-I**) NAACCR (**EUCAIM**)
	2.7 Histological cell type, typically in ICD-O-3 standards	**OSIRIS** (Primary cancer diagnosis present in Tumor pathology event.Morphology Code and Topography Code) **IDEA4RC** (CancerEpisode.HistologySubGroup) **EOSC4Cancer** (CRC Diganosis) **EUCAIM** (PrimaryCancerCondition.HistologyMorphologyBehavior) **ECNR** **ProCAncer-I** (Condition_occurrence.condition_concept_id)	Condition.Neoplasm defined only by histology	ICD-O3 (**OSIRIS, OMOP, EOSC4cancer, EUCAIM, ENCR, ProCAncer-I**) SNOMED (**IDEA4RC, ProCAncer-I**)
	2.8 Menopausal status (for example, for patients with breast cancer)	**EUCAIM** (Observation)	Condition.Menopause finding (finding)	SNOMED (**OMOP, EUCAIM**)
3. Biomarkers	3.1 Biomarker name	**OSIRIS** (Analysis->Biomarker) **BBMRI-ERIC** (Molecular markers)	To be discussed (examples required)	LOINC (**OSIRIS, INCISIVE, AIDAVA, EUCAIM**)

		INCISIVE (CodeSystem the name and the code with LOINC code) **AIDAVA** (Measurment.hasCode) **EUCAIM** (TumorMarkerTest.Code.) **ProCAncer-I** (Measurement.measurement_concept_id)		SNOMED (**AIDAVA, ProCAncer-I**)
	3.2 Biomarker measure	**OSIRIS** (Analysis->Biomarker) **INCISIVE** (Observation resource) **AIDAVA** (Measurment.hasMeasurementMethod) **EUCAIM** (TumorMarkerTest.ValueUnitConcept.) **ProCAncer-I** (Measurement.unit_concept_id)	Measurement	LOINC (**OSIRIS, AIDAVA, EUCAIM**) SNOMED (**OMOP, AIDAVA**) CPT4 (**OMOP**)
	3.3 Biological sample ID	**OSIRIS** (Analysis->Biological Sample) **BBMRI-ERIC** (Sample: SAMPLE_ID) **INCISIVE** (CodeStstem the name and the code with LOINC code) **AIDAVA** (Measurment.hasSample)	Measurement	LOINC (**OSIRIS, OMOP, INCISIVE, EUCAIM**)
4. Treatment	4.1 Line of therapy (derived algorithmically within each cancer type)	**OSIRIS** (Tumor pathology event-> Treatment.LineNumer) **IDEA4RC** (Derived from Treatment entities) **INCISIVE** (Procedure resource) **AIDAVA** (Procedure.hasIntent) **EOSC4Cancer** (Procedure_occurence) **EUCAIM** (Procedure) **ProCAncer-I** (Through the Episode and Episode_Event)	Procedure	SNOMED (**OMOP, INCISIVE, EUCAIM**) OMOP Episode (**ProCAncer-I**)
	4.2 Anti-cancer treatment name, including	**OSIRIS** (Tumor pathology event-> Treatment.Type) **IDEA4RC** (Systemic treatment, radiotherapy and surgery have its own representation, and everyone	Procedure	SNOMED (**OMOP, IDEA4RC, INCISIVE, AIDAVA, EUCAIM,**

systemic treatment and supportive therapy	of this treatment can be also palliative (this is a specific characterization for head and neck and sarcoma)) **INCISIVE** (Procedure resource) **AIDAVA** (Procedure.hasCode) **EOSC4Cancer** (Treatment Episode: CAcner Drug Treatment) **EUCAIM** (Procedure) **ProCAncer-I** (Procedure.Procedure_concept_id) **ONCOFAIR** (CarePlan.identifier/CarePlan/title)		**ProCAncer-I, ONCOFAIR**)
4.3 Molecule generic name	**OSIRIS** (Tumor pathology event-> Treatment -> Drug.name) **IDEA4RC** (SystemicTreatment) **INCISIVE** (Procedure resource) **AIDAVA** (Procedure.usingSubstance.Substance.hasGenericName) **EUCAIM** (Medication Administration) **ONCOFAIR** (Medication.code)	Drug	ATC (**OSIRIS, INCISIVE, EUCAIM**) RxNorm (**OMOP, EUCAIM**) OMOP (**IDEA4RC**) SOMEND (**ONCOFAIR**)
4.4 Start date for drug treatment	**OSIRIS** (Tumor pathology event-> Treatment.StartDate), IDEA4RC (Present in SystemicTreatment_StartDateSystemicTreatment) **BBMRI-ERIC** (Pharmacotherapy: PHARMACOTHERAPY_START_RELATIVE) **INCISIVE** (Present in MedicationAdministration date)	Drug.Drug_exposure_start_date	ISO8601 (**OSIRIS**) SNOMED (**OMOP**)

	PanCareSurPass (in MedicationAdministration.effectivePeriod.start (required)) **AIDAVA** (Procedure.haStartDateTime) **EOSC4Cancer** (Condition Start Date) **EUCAIM** (CancerRelatedMedicationAdministration.effectiveStartDate) **ONCOFAIR** (MedicationAdministration.occurence/MedicationRequest.effectiveDosePeriod)		
4.5 Treatment dose	**IDEA4RC** (Derived from SystemicTreatment.regimen (clinical knowledge)) **BBMRI-ERIC** (Pharmacotherapy: PHARMACOTHERAPY_SCHEME, PHARMACOTHERAPY_SCHEME_DESCRIPTION) **INCISIVE** (MedicationAdministration resource) **PanCareSurPass** (MedicationAdministration.dosage.dose (required) and Observation.value:valueQuantity (required)) **AIDAVA** (Procedure.usingSubstance.Substance.hasQuantity) **EOSC4Cancer** (drug_exposure) **ONCOFAIR** (MedicationAdministration.dosage/MedicationRequest.doseAndRate)	Measurement Calculated usind drug_exposure in combination with drug_strength	RxNorm (**OMOP**)
4.6 End date for drug treatment	**OSIRIS** (Tumor pathology event-> Treatment.EndDate)	Drug.Drug_exposure_end_date	ISO8601 (**OSIRIS**) SNOMED (**OMOP**)

	IDEA4RC (SystemicTreatment_EndDateSystemicTreatment **BBMRI-ERIC** (Pharmacotherapy: PHARMACOTHERAPY_END_RELATIVE) **INCISIVE** (Adjuvant treatment, Neoadjuvant treatment in a Observation.partOf MedicationAdministration) **PanCareSurPass** (MedicationAdministration.effectivePeriod.end (requires)) **AIDAVA** (Procedure.haEndDateTime) **EOSC4Cancer** (Condition End Date) **EUCAIM** (CancerRelatedMedicationAdministration.effectiveEndDate) **ONCOFAIR** (MedicationAdministration.occurence/MedicationRequest.effectiveDosePeriod)		
4.7 Radiotherapy type	**OSIRIS** (Tumor pathology event-> Treatment.Type) **IDEA4RC** (Radiotherapy.setting) **INCISIVE** (Procedure resource) **PanCareSurPass** (Present in Procedure.code (required)) **AIDAVA** (Consent.hasStatusCode) **EOSC4Cancer** (Treatment Episode: Cancer Radiotherapy) **EUCAIM** (RadiotherapyCourseSummary.technique) **ProCAncer-I** (Procedure.procedure_concept_id)	Procedure	SNOMED (**OMOP, IDEA4RC, AIDAVA, EUCAIM, ProCAncer-I**)

4.8 Radiotherapy start date	**OSIRIS** **IDEA4RC** (Radiotherapy.startDate) **BBMRI-ERIC** (Present in Radiation therapy: RADIATION_THERAPY_START_RELATIVE) **INCISIVE** (Procedure resource) **PanCareSurPass** (Procedure.performedPeriod.start (required)) **AIDAVA** (RadiotherapyProcedure.hasCode) **EOSC4Cancer** (Condition Start Date) **EUCAIM** (RadiotherapyCourseSummary.startDate)	Procedure.procedure_date	SNOMED (**OMOP**)
4.9 Radiotherapy dose	**OSIRIS** (Radiotherapy module Table.TotalDosis) **IDEA4RC** (Three variables in Radiotherapy (one will probably be removed): Total Dose (TD) Gy, Fraction Size (FS), Number of fractions)) **INCISIVE** (MedicationAdministration resource) **PanCareSurPass** (Procedure.valueQuantity.value (required)) **AIDAVA** (RadiotherapyProcedure.usingSubstance.Substance.hasQuantity) **EUCAIM** (RadiotherapyCourseSummary.totalDoseDelivered, RadiotherapyCourseSummary.doseDeliveredtoVolume, RadiotherapyCourseSummary.numberOfSessions, RadiotherapyCourseSummary.fractionsDelivered)	Measurement	SNOMED (**IDEA4RC**)

	ProCAncer-I (Measurement.measurement_concept_id and Measurement.unit_concept_id Measurement.value_as_number)		
4.10 Radiotherapy end date	**OSIRIS** **IDEA4RC** (Radiotherapy.endDate) **BBMRI-ERIC** (Radiation therapy: RADIATION_THERAPY_END_RELATIVE) **INCISIVE** (Procedure resource) **PanCareSurPass** (Procedure.performedPeriod.end (required)) **AIDAVA** (RadiotherapyProcedure.hasEndDateTime) **EOSC4Cancer** (Condition End Date) **EUCAIM** (RadiotherapyCourseSummary.endDate)	Procedure.procedure_end_date	SNOMED (**OMOP**)
4.11 Surgery type	**OSIRIS** (Tumor pathology event-> Treatment.Type) IDEA4RC (Present in Surgery.type) **BBMRI-ERIC** (Surgery: SURGERY_TYPE) **INCISIVE** (Procedure resource) **PanCareSurPass** (Procedure.text (optional)) **AIDAVA** (Procedure.haCode) **EOSC4Cancer** (CRC Surgery) **EUCAIM** (CancerRelatedSurgicalProcedure.code) **ECNR** **ProCAncer-I** (Procedure.procedure_concept_id)	Procedure	SNOMED (**OMOP, IDEA4RC, AIDAVA, EOSC4cancer, EUCAIM, ProCAncer-I**) CPT4 (**EUCAIM**)
4.12 Surgery date	**OSIRIS** (Tumor pathology event-> Treatment.StartDate) **IDEA4RC** (Surgery.DateofSurgery)	Procedure.procedure_date	SNOMED (**OMOP**)

		BBMRI-ERIC (Surgery: SURGERY_START_RELATIVE) **INCISIVE** (Procedure resource) **PanCareSurPass** (Procedure.performedDateTime (required)) **AIDAVA** (Procedure.haStartDateTime) **EOSC4Cancer** (Mesurement Date) **EUCAIM** (CancerRelatedSurgicalProcedure.effective)		
	4.13 Participation in clinical trial	**OSIRIS** (Tumor pathology event-> Treatment.Clinical TrialId) **BBMRI-ERIC** (Patient Data: CLINICAL_STUDY_PARTICIPANT) **INCISIVE** (Present in the user interface) **PanCareSurPass** (Procedure.instantiatesCanonical (optional))	Observation: Enrollment in clinical trial - Concept ID 46271379	SNOMED (**OMOP**)
	4.14 Date of trial consent	**INCISIVE** (Present in the user interface)	Observation	
5. Outcomes	5.1 Date of death, at any location	**OSIRIS** (Patient) **AIDAVA** (DeathDate) **EOSC4Cancer** (Date of Death) **EUCAIM** (CancerPatient.DateofLastContact if CancerPatient.deceased) **ECNR**	Death.death_date	SNOMED (**OMOP**) LOINC (**EUCAIM**)
	5.2 Time to next treatment (derived from	**INCISIVE** (Procedure resource)	To be discussed (no future events)	

treatment start dates)			
5.3 Metastasis presence/absence	**IDEA4RC** (Stage.Is metastatic (associated to EpisodeEvent) **BBMRI-ERIC** (TIME_OF_RECURRENCE_RELATIVE) **INCISIVE** (Observation resource) **PanCareSurPass** (Condition.conformsTo (optional)) **AIDAVA** (ProblemCondition) **EOSC4Cancer** (Disease Extent: Metastatic Disease) **EUCAIM** (SecondaryCancerCondition.code) **ProCAncer-I** (Episode and Episode_Event)	Episode Measurement and in Episode as Disease Extent	OMOP Episode (**OMOP, ProCAncer-I**) SNOMED (**AIDAVA, EUCAIM**) IDC-10 (**AIDAVA**)
5.4 Metastasis location	**IDEA4RC** (Binary flags in Stage (associated to EpisodeEvent)) **INCISIVE** (Observation resource) **PanCareSurPass** (Condition.bodySite.coding (optional)) **AIDAVA** (ProblemCondition.hasBodySite) **EUCAIM** (SecondaryCancerCondition.bodySite) **ProCAncer-I** (Measurement and Condition)	Measurement	OMOP Cancer Modifier (**OMOP**, **IDEA4RC, EUCAIM, ProCAncer-I**) ICD-O3 (**PanCareSurPass**)
5.5 Date of clinical visits (with cancer related visits separated from other visits)	**OSIRIS** **INCISIVE** (Procedure resource) **AIDAVA** (HealthcareEncounter.hasStartDateTime)	Visit.start_date, Visit.end.date	SNOMED (**OMOP**)
5.6 Vital status (derived from	**OSIRIS** **IDEA4RC** (Derived from PatientFollowUp.Status at last follow-up)	Present and can be derived from DEATH table with	SNOMED (**OMOP, EOSC4cancer**)

visits or death linkage)	**BBMRI-ERIC** (VITAL_STATUS, VITAL_STATUS_TIMESTAMP) **AIDAVA** (ProblemCondition) **EOSC4Cancer** (Cause of Death) **EUCAIM** (CancerPatient.deceased) **ECNR**	DEATH.cause_concept_id	
5.7 Extent of debulking surgery (for example, for patients with gynecological cancer)	**INCISIVE** (Procedure resource)	Measurement	SNOMED (**OMOP**)

Table 2 The results of the comparison between the data concepts defined in MEDOC and the ones adopted in each project. Records in light blue correspond to the first phase survey while purple ones to the second phase.

Project	Project objectives	Primary or secondary use of data?	Use cases for the project	Is the data model proprietary?	Is the data model mapped to a standard one (FHIR, OMOP, openEHR etc)?	Are there any redundancy in the model to facilitate the query for the clinicians?
MEDOC	- re-use of secondary data for natural history and outcome studies.	Secondary	Solid tumours, ongoing studies on breast, lung and colorectal	No	OMOP	

INCISIVE		Secondary	Breast, Colorectal, Prostate and Lung cancers	No	FHIR	No
IDEA4RC	- Secondary use of data for natural history, prognostic and predictive studies and studies on treatment effectiveness and quality of care.	Secondary	- Incidence of skeletal metastases (after diagnosis) in patients with solitary fibrous tumours (SFT) in general and by site of primary SFT (e.g., meningeal versus extra meningeal etc.) - Validation of the prognostic significance of neutrophils/lymphocytes ratio (NLR) and prognostic index combining serological and inflammatory factors (PISIF) in primary retroperitoneal sarcomas - Identification of predictors of outcome after surgical treatment (with respect to both short term morbidity, survival, recurrences and quality of life) in H&N cancers. - Assessment of the outcomes (overall survival, disease free survival) of sino-nasal cancer patients treated with induction chemotherapy. - Induction chemotherapy is a treatment approach where chemotherapy is administered before the main treatment, such as surgery or radiation therapy." - Assessment of the role of photon and proton-based radiotherapy on the outcomes (overall survival, disease free survival) of low and intermediate grade	No	Both FHIR and OMOP	Yes

			mucoepidermoid cancers of salivary gland. - Assessment of the outcomes (overall survival, disease free survival) of locally advanced salivary gland cancers treated with surgery + radiotherapy +/- chemotherapy - Treatment of first progression in head and neck cancers			
OSIRIS	OSIRIS (Validated) : Oncology data data from clinical trial/e-FORM to a second reuse for oncology research OSIRIS_RWD (Unvalidated) : Oncology data from EHR to a second reuse for oncology research	Secondary	Precision medicine (high granularity for pathology and genomic data)		FHIR	
ONCOFAIR		Secondary	Prescription and administration for oncological treatement			

PanCareSurPass	To improve Person-Centered Survivorship Care	Secondary	Multi-country Implementation Study: to implement the digital SurPass version 2.0 in clinics (with three different scenarios) across Austria, Belgium, Germany, Italy, Lithuania, and Spain 2) Cost-analysis: to assess the suitability and costs of using the digital SurPass in different contexts			
EUCAIM		Secondary	Aims to build a distributed, federated Atlas of Cancer Images, including both common and rare types of cancer, for the development, training, and benchmarking of AI-powered tools for cancer management.	No	OMOP and FHIR (mCODE).	
ProCAncer-I		Secondary	1) Detection of prostate cancer with high accuracy both in peripheral and transitional zones to identify which men have cancer and those with no cancer. 2) Characterization of cancer according to its biological aggressiveness into clinically significant and non-significant disease. 3) Identification of patients with metastatic prostate cancer as early as possible among cases with high-risk PCa. 4) Radiologic – Histopathologic correlation to provide biology-based validation of AI models to compare side by side pathologic data with AI results to improve	No	OMOP	

			understanding of the features that AI models are making use to reach specific decisions. 5) Prediction of the risk of disease recurrence after radical prostatectomy, based on imaging data and AI techniques. 6) Prediction of treatment response in case of radiation therapy, assessing the risk of disease recurrence to promptly adjust therapeutic strategy at an early stage and avoid patient discomfort and non-optimal distribution of medical resources. UC6 is similar to UC5, but refers to radiotherapy recurrence and it will help radiation oncologist to tailor treatments; 7) Prediction of post radical prostatectomy and/or radiation-induced urinary toxicity, in order to consider additional or alternative measures to alleviate therapy-induced undesired effects. 8) AI-powered patient stratification for enrolment in Active Surveillance programs, to develop a more efficient patient stratification program based on AI decision-making from MRI lesion phenotype. 9) Prediction of the best option for patients needing			

			treatment ensuring the lowest possible side effects/toxicity.			
EHDS2Pilot	Develop and deploy an IT infrastructure for setting up a network of nodes and delivering basic central services jointly with the European Commission, such as common data discovery, common data access application, and allowing data usage (and possibly data transfer).	Secondary	Infectious disease surveillance (AMR): Demonstrate the feasibility of using the EHDS to carry out infectious disease surveillance, focusing on antimicrobial resistance. Thrombosis in COVID-19 patients: Foster a better understanding of the risks of thrombosis in COVID-19 patients. Covid-19 testing, vaccination and hospitalisation: Compare COVID-19 testing, vaccination and hospitalisation between the general population and vulnerable subpopulations. Cardiometabolic diseases: Compare care pathways for cardiometabolic diseases in European countries and build prediction models, using artificial intelligence. Colorectal cancer: Mobilise and chain clinical and genomic data to enhance our understanding of colorectal cancer.			
ENCR		Secondary	European cancer statistics based on population-based cancer registry data	Yes, but using standardize	OMOP	

				d vocabularies		
German Medical Informatics Initiative / German Cancer Registries	FHIR profiling of the German Medical Informatics Inititaives Extension Module based on the official core dataset of all German Cancer Registries	First version secondary Later versions elso primary data collection.	Making Data of Cancer Registries available for federated research			
AIDAVA		-	Demonstrate that we can generate automatically a federated interoperable BC registry from curated high-quality interoperable patient records	-	-	-
EOSC4cancer		Secondary	Task 4.1: Cancer risk identification and prevention by linking environmental data to cancer registry data Task 4.2: Data-driven optimisation of cancer screening programs Task 4.3: Data-driven treatment selection for localised tumours with multiple patient-derived data types Task 4.4: Data-driven treatment selection for localised	No	OMOP	

			tumour: improving the treatment of colorectal cancer by the inclusion of circulating DNA information Task 4.5: Connecting omics data from multiple sources to a Clinical Decision Support System for precision treatment of metastatic colorectal cancer			

Table 3 The summarization of the result obtained from the sheet "Projects Contracts"

Project or initiative	Data model link
OSIRIS	https://drive.google.com/file/d/17_lT8ozj1krW8ygnCO2UVSS-hbllYbud/view?usp=drive_link
IDEA4RC	https://docs.google.com/spreadsheets/d/1ANErBpHQAW6ngn1kq-a7rPpeTosG-z2PHnwfUT6IUKI/edit?gid=592552349#gid=592552349
PanCareSurPass	https://docs.google.com/presentation/d/1l8aIZMGfBiJV48jgNZNhp71p_D1s2TLl/edit#slide=id.p1 https://build.fhir.org/ig/hl7-eu/pcsp/logicalModel.html
1+MG	https://drive.google.com/file/d/1-gTS46pqGHwPh8v_DofbHTzPJzTxoBVd/view?usp=drive_link DETAILS https://drive.google.com/drive/folders/15kLJHB61PKUaGkodS1jaN4kpd9ZuO5Yq
EHDS2Pilot (BBMRI CRC)	https://www.bbmri-eric.eu/wp-content/uploads/D2.4_rev3.pdf (from page 12)
ENCR/IKNL	https://plugin.healthcare/fhir/artifacts.html#structures-logical-models
ONCOFAIR	https://oncofair.github.io/FHIR-ImplementationGuide-Article/main/ig/mapping.html

Table 4 List of data model representations shared by the projects or initiatives with HL7 Europe

Passing to the information collected in the sheet "ProjectContacts" of the "Comparison between Models v1" reported in Table 3 and Table 5, it is possible to summarize that as the following:

- The objectives and the user cases of the involved projects and initiatives are different, but all are based on the secondary data use. This confirms that the European Cancer Common Data Model should be designed to support secondary data use.
- The shared data models are not proprietary
- They are mapped in FHIR and/or OMOP
- Only IDEA4RC seems to have the need to use redundancy.

Project	FHIR IG
OSIRIS	https://build.fhir.org/ig/InstitutNationalduCancer/ImplementationGuide_OsirisFHIR/
IDEA4RC	https://build.fhir.org/ig/hl7-eu/idea4rc/
INCISIVE	https://fhir.incisive-project.eu/index.html https://simplifier.net/INCISIVE/~introduction, https://incisive-project.eu/wp-content/uploads/2023/08/INCISIVE_D3.4_StandardizationSuggestions_v1.0_FinalVersion.pdf
PanCareSurPass	https://hl7.eu/fhir/ig/pcsp/
FLUTE	https://build.fhir.org/ig/hl7-eu/flute/
ONCOFAIR	https://github.com/ONCOFAIR/FHIR-ImplementationGuide-Article
EHDS2Pilot	https://github.com/samply/bbmri-fhir-ig
German Medical Informatics Initiative / German Cancer Registries	https://simplifier.net/guide/mii-ig-modul-onko-2024-de?version=current (German only)
IKNL	https://github.com/IKNL/NkrIG https://simplifier.net/iknl-ncr-ehr-r4 https://simplifier.net/iknl-palga-r4 https://plugin.healthcare/fhir

Table 5 The list of FHIR Implementation Guides implemented by each project or initiative. Records in light blue correspond to the first phase survey while purple ones to the second phase.

Finally, the sheet "New variables" was completed only by IDEA4RC and ENCR (Figure 43). ENCR highlighted an aspect in alignment with the output derived from the WP2 activities: the cancer patients are not always treated in single hospitals, and it is fundamental to reconstruct the patient journey and the history of the disease.

4.2.5 The European Cancer Common Conceptual Model

As previously indicated in chapter 3.2.5**Errore. L'origine riferimento non è stata trovata.**, all the results of the survey, especially the ones reported in chapters 4.2.1**Errore. L'origine**

riferimento non è stata trovata. and 4.2.4, were adopted to design, with the fundamental help of partners that worked on the IDEA4RC data model (WP2), a preliminary draft of a possible model of the entities and their relationship to be present, discuss and consolidate it during the official event. For the design the diagram of entity and their relationships defined by mCODE was adopted as the starting point. The prepared and discussed entity model is represented in Figure 57.

The list of priority entities on which start to work prepared by elaborating the results of the conducted survey based on MEDOC concepts is presented in Table 6 where it is possible to see that the MEDOC concepts on which start to work should have been: sex, healthcare ID (or other unique identifier), Primary cancer diagnosis and comorbidities, typically in International Classification of Disease standards such as ICD10, ICD9 or ICD-O-3, Surgery type, Data of primary cancer diagnosis, disease stage in a recognized standard such as TNM, start date for drug treatment and end date for drug treatment.

Table 6 also reports the map between the MEDOC concepts and the corresponding entities in the prepared draft model and in the mCODE model prepared to guide the working section, with the fundamental contribution of the WP2.

All the presented material was adopted as stating point for the "Common Cancer Model Track" (Figure 58) that was a great success for HL7 Europe. Considering the importance of the topic, it was the longest track at the event with about 35 participants (7 of them from IDEA4RC consortium) coming from 11 European countries: Italy, Germany, Netherlands, France, Norway, Spain, Switzerland, Slovenia, United Kingdom, Czechia, and Greece. A fundamental aspect was the heterogeneity of attendees: clinical and research doctors, technical experts (FHIR/OMOP, Data modelling). It was the first HL7 multidisciplinary track ever, highlighting the efficient activities performed by HL7 Europe to spread this initiative performed in Task 10.3 and the importance that the topic has not only for IDEA4RC but also for the European community that works on the cancer domain.

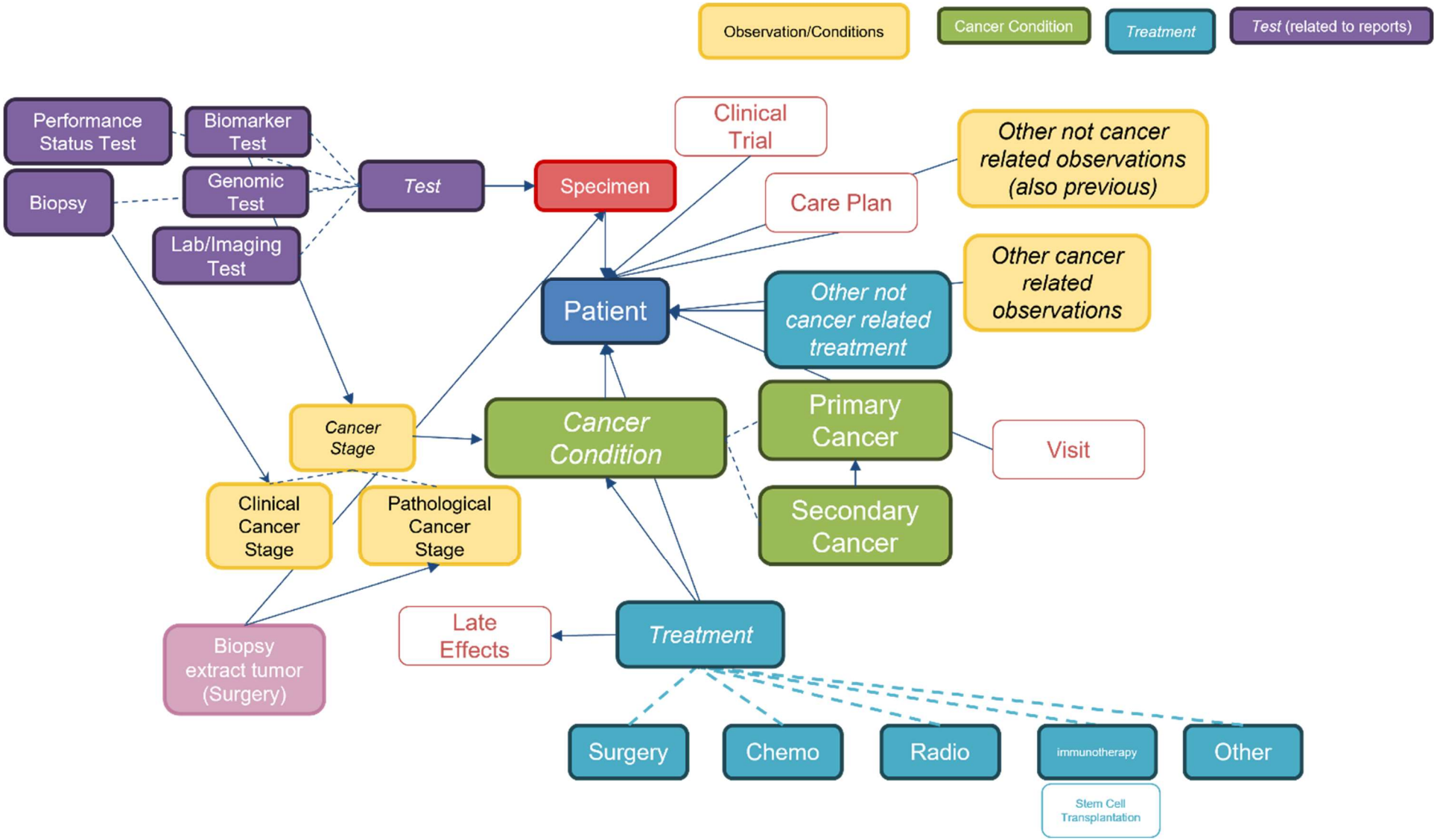


Figure 57 The preliminary draft of a possible model of the entities and their relationship prepared for the "Common Cancer Model Track"

MEDOC Categories	MEDOC concepts	Priority	Entity in the draft model	Entities in mCODE
1. Demographics	1.1 Date of birth (month)	9	Patient	Cancer Patient
	1.2 Sex	12	Patient	Cancer Patient
	1.3 Weight (with timestamp)	3	Patient	Cancer Patient
	1.4 Height	3	Patient	Cancer Patient
	1.5 Healthcare ID (or other unique identifier)	11	Patient	Cancer Patient
	1.6 Legal basis for data processing	4		
2. Clinical phenotype	2.1 Primary cancer diagnosis and comorbidities, typically in International Classification of Disease standards such as ICD10, ICD9 or ICD-O-3	11	PrimaryCancerCondition Other not cancer related observations (also previous)	Comorbidities
	2.2 Charlson comorbidity index (derived from 17 comorbidities in 2.1)	2	Other not cancer related observations (also previous)	Comorbilities
	2.3 Date of primary cancer diagnosis	10	PrimaryCancerCondition	PrimaryCancerCondition
	2.4 Method of primary cancer diagnosis	8	Test related to CancerStage	Test related to CancerStage
	2.5 Performance status (for example, coded by ECOG or Karnofsky standards)	6	Performance Status Test	ECOG Performance Status referring to CancerPatient
	2.6 Disease stage in a recognized standard such as TNM	10	Cancer Stage	CancerStage / TNM Stage Group
	2.7 Histological cell type, typically in ICD-O-3 standards	7	PrimaryCancerCondition	Histologic Behavior and Type / Tumor Morphology Report

				/ Primary Cancer Condition
	2.8 Menopausal status (for example, for patients with breast cancer)	2	Other not cancer related observations (also previous)	-
3. Biomarkers	3.1 Biomarker name	7	Biomarker Test	Tumor Marker Test
	3.2 Biomarker measure	6	Biomarker Test	Tumor Marker Test
	3.3 Biological sample ID	5	Biomarker Test	Tumor Marker Test
4. Treatment	4.1 Line of therapy (derived algorithmically within each cancer type)	8	Treatment	Treatments related to CancerCondition
	4.2 Anti-cancer treatment name, including systemic treatment and supportive therapy	9	Treatment	Cancer-Related Medication Request / Administration
	4.3 Molecule generic name	6	Treatment	Cancer-Related Medication Request / Administration
	4.4 Start date for drug treatment	10	Treatment	Cancer-Related Medication Request / Administration
	4.5 Treatment dose	8	Treatment	Cancer-Related Medication Request / Administration
	4.6 End date for drug treatment	10	Treatment	Cancer-Related Medication Request / Administration
	4.7 Radiotherapy type	9	Treatment	Radiotherapy Course Summary
	4.8 Radiotherapy start date	9	Treatment	Radiotherapy Course Summary
	4.9 Radiotherapy dose	8	Treatment	Radiotherapy Course Summary
	4.10 Radiotherapy end date	9	Treatment	Radiotherapy Course Summary

	4.11 Surgery type	11	Treatment	Cancer-Related Surgical Procedure
	4.12 Surgery date	8	Treatment	Cancer-Related Surgical Procedure
	4.13 Participation in clinical trial	5	ClinicalTrial	-
	4.14 Date of trial consent	2	ClinicalTrial	-
5. Outcomes	5.1 Date of death, at any location	6	Patient	Cancer Patient
	5.2 Time to next treatment (derived from treatment start dates)	2	CarePlan	Treatment
	5.3 Metastasis presence/absence	9	Secondary Cancer Condition??	Secondary Cancer Condition
	5.4 Metastasis location	7	Secondary Cancer Condition??	Secondary Cancer Condition
	5.5 Date of clinical visits (with cancer related visits separated from other visits)	4	Visit	-
	5.6 Vital status (derived from visits or death linkage)	9	Patient	OtherObservations ?? Derived from Death Date
	5.7 Extent of debulking surgery (for example, for patients with gynecological cancer)	2	Surgery	Cancer-Related Surgical Procedure

Table 6 Priority list of MEDOC concepts and map to the entities of the proposed draft model and the ones of mCODE

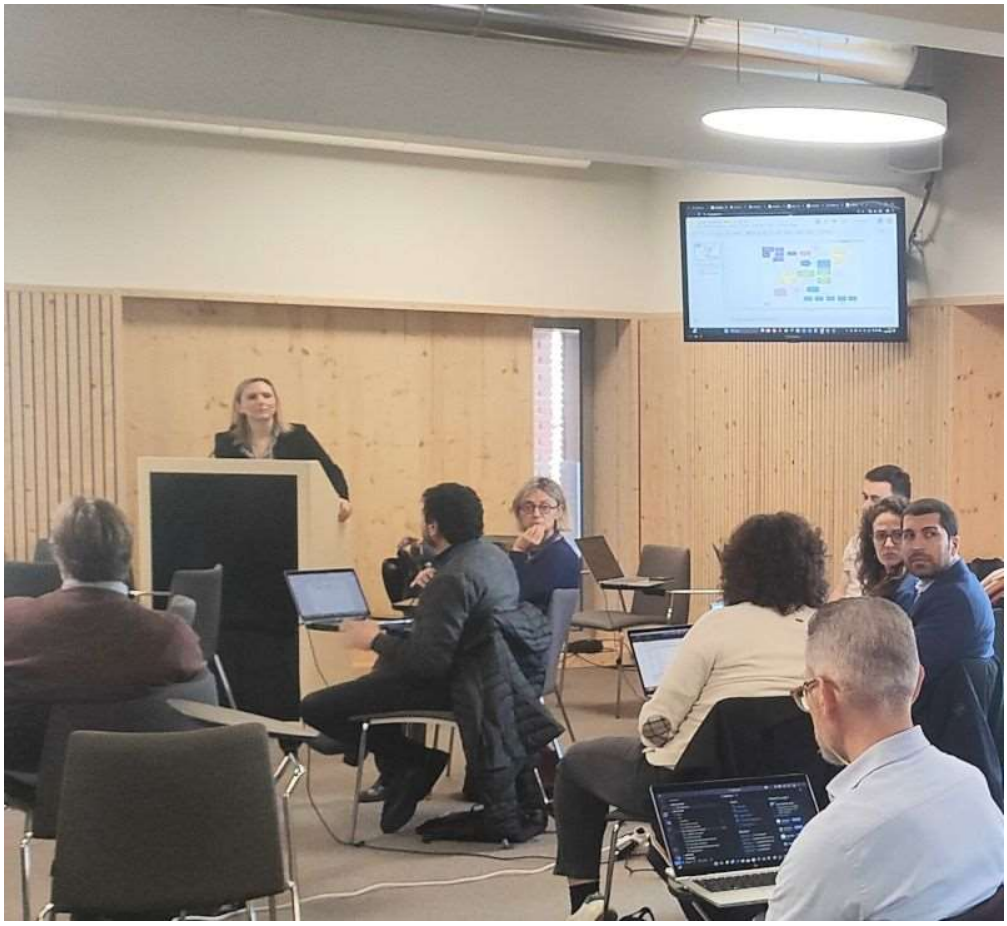


Figure 58 The WG in action during the "Common Cancer Model Track"gggf

During the track the cancer patient's journey defined in IDEA4RC was presented. During the discussion the WG agreed that it should be better to call it cancer journey because the focus is not the patient (as it is in mCODE) but is the cancer condition. This was the first relevant results, highlighting that the model should be based on the cancer condition rather than the patient.

After this agreement, the activities moved on the entity model as indicated in chapter 3.2.5**Errore. L'origine riferimento non è stata trovata.**. As result of the track, the WG worked on the first draft of the European Cancer Common Conceptual Model represented in Figure 60 and Figure 61 where it is possible to the relevant entities that the WG decided to be the relevant ones as the CancerConditin, Clinical Stage, Disease Extent, Visit, Lab Report and Treatment.

The track was concluded with the decision to organized monthly calls for March and April to continue the discussion on treatments and on metastatic cancer. The detail of the track and its results will be published in the "2025-02-12 Q1-Q2-Q3-Q4/2025-02-13 Q2 - Common Cancer

Model Track"[77] page of HL7's Confluence (Figure 59) to allow to spread and share them with everyone who wants to collaborate.

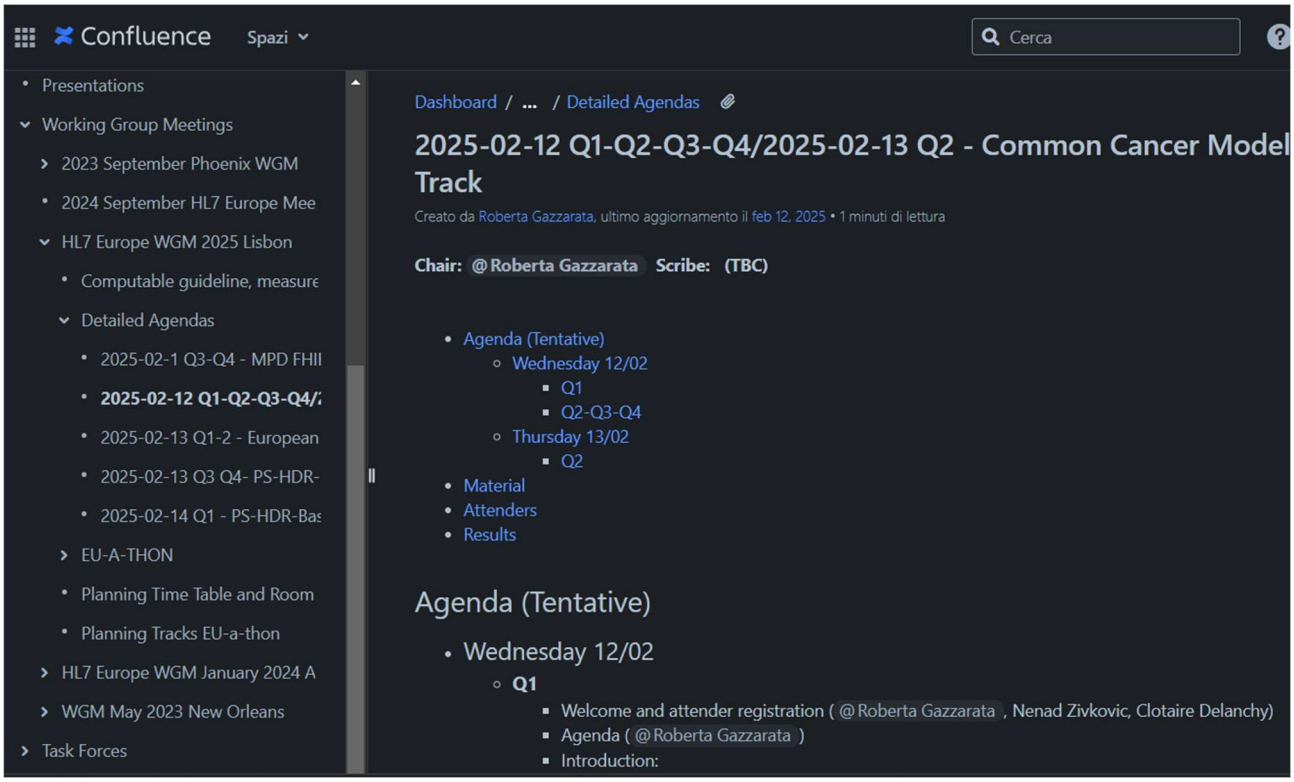


Figure 59 The "European Cancer Mission Track" page in the HL7's Confluence

[77] https://confluence.hl7.org/spaces/HEU/pages/308976299/2025-02-12+Q1-Q2-Q3-Q4+2025-02-13+Q2+-+Common+Cancer+Model+Track

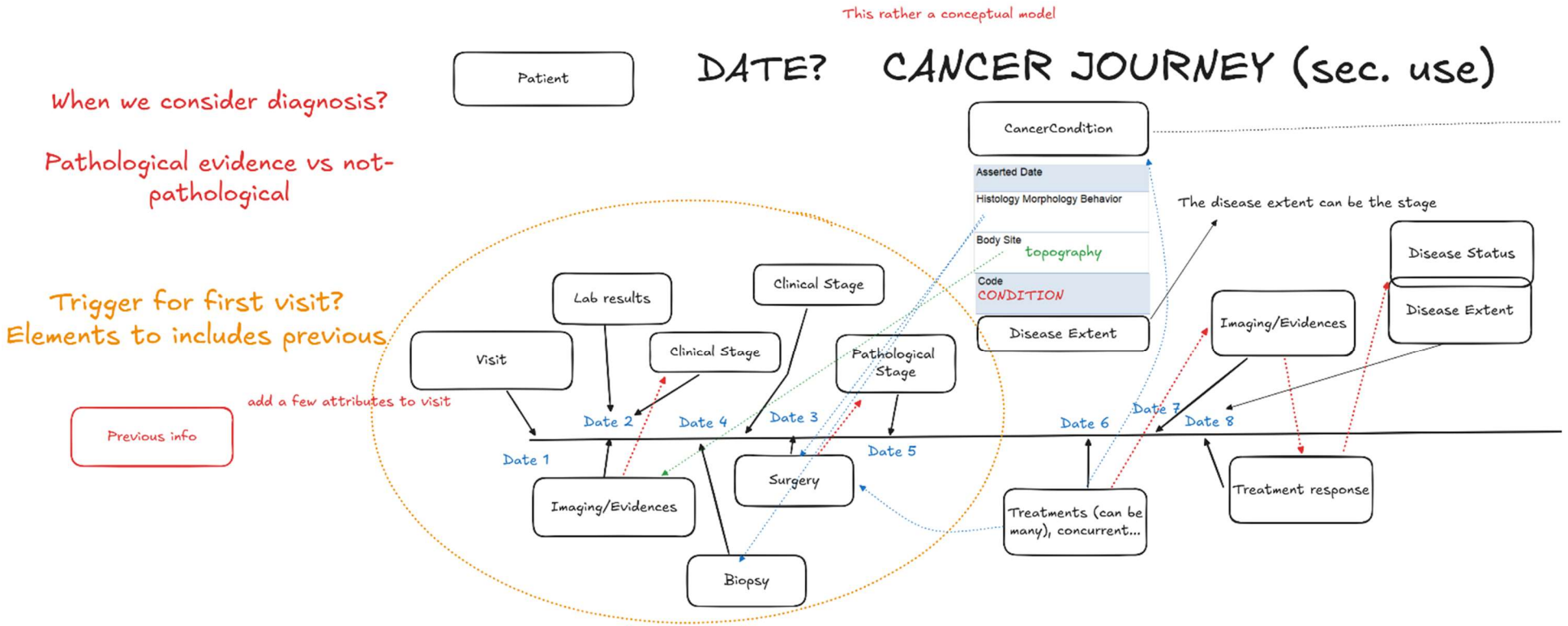


Figure 60 The conceptual model developed (part 1)

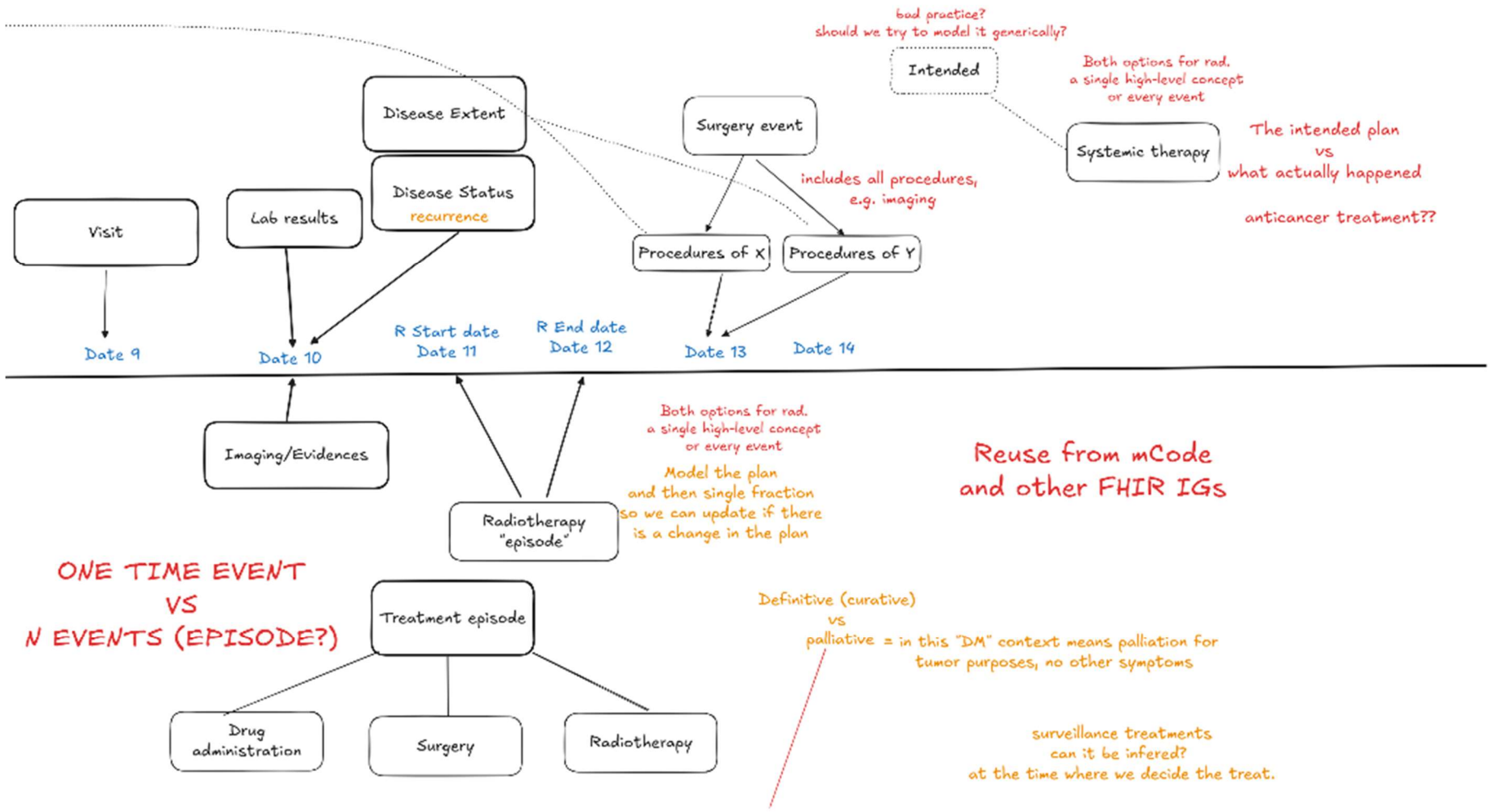


Figure 61 The conceptual model developed (part 2)

5 CONCLUSIONS

This document, part of the Task T10.3 "Ecosystem standard", reported an overview of the European and international context on cancer data ecosystem, the activities performed from month 13 to month 30 of IDEA4RC project on the analysis of standards for a Rare Cancer Data Ecosystem and on standards relevant for the enlarged ecosystem achieved as result of Task 10.3n, and on the first steps taken to create an European Cancer Common Data Model.

Chapter 2 proposed the European and international context on cancer data ecosystem. At European level, the presentation started with the Europe's Beating Cancer plan and the European Health Data Space (EHDS), followed by the European EHR Exchange Format (EEHRxF) and the related X-eHealth project. Then the following projects were presented:

- the European Commission funded projects FLUTE, PanCareSurPass, INCISIVE, EUCAIM, Genomic Data Infrastructure (GDI), EHDS2Pilot
- the cluster Cancer Survivorship - AI for Well-being (CS-AIW)
- the national initiatives OSIRIS (French), German Cancer Registries / German Medical Informatics Initiative (MII) Oncology Module (Germany), The Netherlands Comprehensive Cancer Organisation (IKNL) (Netherlands)
- the European networks and research infrastructure EURACAN, European Network of Cancer registry (ENCR), BBMRI-ERIC, DIGICORE.

An introduction of the European project, part of DIGICORE, DIGItal Oncology Network for Europe (DIGIONE) and its Minimal Essential Description of Cancer (MEDOC) was also provided for the relevant impact they had on the performed activities reported in this document. At international level, The HL7 FHIR Accelerators Vulcan and CodeX were presented as programs designed to support communities and collaborative groups across the global healthcare spectrum in creating and adopting high-quality FHIR Implementation Guides, promoting health data interoperability. The focus was on their products and results most relevant for the IDEA4RC purposed: the "FHIR to OMOP" project (Vulcan FHIR Accelerator), the "OHDSI OMOP + FHIR" Working Group, the mCODE (minimal Common Oncology Data Elements) FHIR Implementation Guide (CodeX FHIR Accelerator).

Chapters 3and 4**Errore. L'origine riferimento non è stata trovata.** described respectively the methodology adopted and the results achieved in the activities performed from month 13 to

month 30 of the IDEA4RC project. The activities were grouped into two main categories: the first ones based on the collaboration and participation to the HL7 Int initiatives Vulcan (FHIR to OMOP), and Codex (mCODE), and the second ones performed at European level with the HL7 Europe Working Group "European Cancer Mission".

In detail, chapter 3.1 described how HL7 Europe interacted and collaborated with the "OHDSI OMOP + FHIR" Working Group of the Vulcan FHIR Accelerator "FHIR to OMOP" project and with Codex FHIR Accelerator on mCODE FHIR Implementation Guide, by participating to the HL7 International Working Group Meetings (WGMs) and adopting the tools provided by HL7 (chat.fhir.org, HL7's Confluence, and HL7's Jira). Chapter 4.1 reported the results obtained in this first phase of the activities starting from sharing with IDEA4RC partners new knowledge achieved as CAMP FHIR tool for then pass to solving technical points to finish with updating the mCODE FHIR IG with feedback collected within IDEA4RC.

The results of this first phase represented a great example and a guideline of how European Commission funded projects collaborate with different initiatives as working groups and technical community of different SDO (as HL7 FHIR Community) to enrich the standard specifications with the experience and the feedback derived by European project as IDEA4RC.

The second period, as explained in detail in chapter 3.2**Errore. L'origine riferimento non è stata trovata.**, was focused on the improvement of primary and secondary cancer data use in the context of the EHDS. The chapter described in detail all the steps which HL7 Europe performed to create a preliminary draft of a European Cancer Common Data Model that then could be used for creating a specific HL7 Europe Cancer Common FHIR Implementation Guide for the cancer domain. The activities started with the proposed and accepted "European Cancer Mission Track" at the first HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024" that was hosted in Athens in January 2024. This first event was the kick-off for a 2-phase survey to compare data models, adopted by each involved European and national project and initiative, basing on the MEDOC concepts. The intent was to understand what are the most adopted MEDOC concepts on which to start to work for the creation of the European Cancer Common Data Model and the corresponding European Cancer FHIR Implementation Guide and collect information useful to start to work and the data model. To extend this survey to other possible interested participants, the workshop "Towards a European Cancer Minimum Data Model and HL7 Europe Cancer Common FHIR Implementation Guide in the EHDS" was proposed, organized and hosted by the MIE conference at Athens in August 2024. At the end of 2024,

considering the promising results on this topic, thought that it was the wright time to have a face to face working session to start to design the European Cancer Data Model, planning the "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025" that was hosted in Lisbon on February 2025 and inviting all the participants of the survey to join bringing their experience maturated in their specific activities.

The detailed results, and their discussion, of each step of this second phase are presented in chapter **Errore. L'origine riferimento non è stata trovata.**. The relevant results can be summarized as the following.

First, there is a general interest and intent from European and national projects and initiatives to collaborate in the definition of a possible European Cancer Common Data Model. It was proved by the great response that HL7 Europe received to collaborate in each step of the activities starting from the questionnaire compilation to prepare the "European Cancer Mission Track" at the "HL7 Europe Working Group Meeting and HL7 FHIR Marathon 2024" to get to the great participation to the "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025" that was hosted in Lisbon in February 2025.

The great collaboration obtained highlighted the need for a European Cancer Common Data Model. In fact, most of the involved projects or initiatives have implemented a specific common data model for its specific needs and requirements but there is some data which is common among all the projects and initiatives. It was clear to all the involved participants that to achieve real interoperability it is necessary to focus on the common data adopted in every project and initiative to try to harmonize them by defining a European Cancer Common Data Model that can be used by the different SDOs to implement specific standards. This aspect was confirmed also by the European Federation of Medical Informatics (EFMI), which during the workshop "Towards a European Cancer Minimum Data Model and HL7 Europe Cancer Common FHIR Implementation Guide in the EHDS" at MIE conference welcomed with enthusiasm this initiative, proving the interest of the activity and motivating to continue.

The conducted survey highlighted important aspects:

- The data models adopted in the different involved projects or initiatives are very heterogeneous. This is understandable because each project or initiative had put its effort into defining a data model that can satisfy the specific needs to achieve its goals.

This confirms that an additional effort, based on the experience of each project is needed to create a European Cancer Common Data Model that could help to facilitate the data sharing among the projects.

- It is difficult to know if the data are available in the EHR of the involved care facilities.
- The objectives of the involved projects and initiatives are different, but all are based on the secondary data use. This confirms that the European Cancer Common Data Model should be designed to support secondary data use.
- There are several MEDOC concepts that are adopted by all (or almost all) the involved projects and initiatives (e.g. Patient sex, Primary Cancer diagnosis and comorbidities) and therefore must be considered with priority in the European Cancer Common Data Model design.
- MEDOC is not a data model, but it is a list of data concepts that can be simple data (e.g. data of birth and sex) or complex data (e.g. Primary cancer diagnosis and comorbidities). In addition, it was created to support the automation of outcomes research studies, and it is based on primary use of data, while the European Cancer Common Data Model should also cover the secondary data use. Therefore, MEDOC is very useful, but it cannot be directly adopted as the basis for the creation of the European Cancer Common Data Model.
- There are some MEDOC concepts that can be mapped in different ways in OMOP (e.g. Data of birth, Sex).
- There are some standard vocabularies that are used in different projects or initiatives therefore can be considered as reference ones of the European Cancer Common Data Model, however an additional deeper analysis should be performed.
- Only IDEA4RC needed to insert redundancy in its data model to facilitate the queries for the clinicians.
- The European Cancer Common Data Model should be able to reconstruct the patient journey and the history of the disease also considering that cancer patients are not always treated in single hospitals.

All these results obtained until the “HL7 Europe Working Group Meeting and EU-a-thon 2025” were partially reported to the participants to the “Common Cancer Model Track” and were adopted to design, with the fundamental help of partners that worked on the IDEA4RC data model (WP2), a preliminary draft of a possible model of the entities and their relationship to be present, discuss and consolidate it during the official event. The draft was prepared starting from the entities and their relationship defined by mCODE focusing on the MEDOC concepts.

A list of priority entities on which to start to work was obtained by the elaboration of the results of the conducted survey based on MEDOC concepts. A map between the MEDOC concepts and the corresponding entities in the prepared draft model and in mCODE model was prepared to guide the working section.

The “Common Cancer Model Track” at the “HL7 Europe Working Group Meeting and EU-a-thon 2025” with about 35 participants coming from 11 European countries was a great success for HL7 Europe. The multidisciplinary WG, starting from the results achieved within IDEA4RC by different WPs, created the first release of the European Cancer Common Conceptual Model.

It is important to mention that the activities performed in the second period required a great effort in coordination with the other projects and initiatives. In several cases a pressing activity was needed to try to get a response (also negative) for the request of collaboration. This for different reasons, but mainly because the projects do not have dedicated budget and time or authorization for this type of collaboration and activities (emerged for example in the call to participate to the 2 HL7 Europe events described in this document). Anyway, in general, they are interested in the topic and want to be updated on the activities that will be performed and the results that will be achieved. These aspects should be considered and reported to the European Commission to improve the cross-projects collaboration.

Considering all these aspects, and the importance of the results obtained with the activities if Task 10.3 reported in this Deliverable, HL7 Europe in 2024 decided to create the first European Initiative for cancer aiming to become the first European HL7 FHIR Accelerator. It was called PHOENIX (Patient Health Oncological Expertise Network for International Exchange) European Initiative and was created to realize a future where seamless data interoperability transforms cancer care and research, enabling breakthroughs that improve patient outcomes and empower collaborative innovation across Europe. In detail, is mission was to play a leading role in the development and adoption of European digital health standards for cancer to support cancer research and care delivery through interoperable data standards, by:

- promoting a unified approach that enhances patient treatment, accelerates research, and closes the loop between care provisioning and research insights
- aligning with and contributing to the European Health Data Space (EHDS) initiatives
- fostering a unified ecosystem where healthcare providers, researchers, and policymakers collaborate to enhance patient outcomes and accelerate innovation.

The PHOENIX Initiative was formally announced at the "HL7 Europe Working Group Meeting and EU-a-thon 2025". In fact, the first project of the PHOENIX Initiative was the "Cancer Common Model Project" with the aim to use the results obtained in this period of activities in IDEA4RC and reported in this document to continue and coordinate the activities to achieve to create a first release of the European Cancer Common Data Model and of a specific HL7 Europe Cancer Common FHIR Implementation Guide in 2025.

In detail, the scope as mentioned will be to define a minimal, extensible, non-exhaustive European cancer data model that will be

- agnostic to the type of cancer, to be
- usable across different use cases
- leveraging on the experiences of the European projects working with primary and secondary usage
- taking into account the availability and usability of reliable data in the EHR systems

and its HL7 FHIR representation in the form of a HL7 FHIR Implementation Guide. As principles, as decided with the IDEA4RC coordinators,

- the activity should be inclusive and cross-cutting, that means transversal to different cancer domains and purposes of use and to consider the needs of different communities
- the approach should be incremental, starting with a minimal core set and then extending / enhancing the model in following iterations
- the dataset should be designed for facilitating the representation with HL7 FHIR.

The first steps of the PHOENIX initiatives will be the organization of the monthly calls for March and April proposed during the "Common Cancer Model Track" at the "HL7 Europe Working Group Meeting and EU-a-thon 2025" to continue the activities on the European Cancer Common Conceptual Data Model. In the meantime, other 2 WGs will be created to work on the European Cancer Common Logical Data Model starting from the conceptual one and on the HL7 Europe Cancer Common FHIR Implementation Guide. The results achieved in the next 2 months, so the conceptual model, the draft of logical model and the draft of the implementation guide, will be presented and discussed at international level (not only European) during a dedicated cancer track at the "HL7 International 2025 May Working Group Meeting & HL7 FHIR

Connectathon" that will be hosted in Madrid in May 2025. During this event HL7 Europe will ask to mCODE group of the CodeX FHIR Accellerator to collaborate to create an International Cancer Common Data Model and a possible HL7 International Cancer Common FHIR Implementation Guide.