*Digital Health and Informatics Innovations for Sustainable Health Care Systems*
*J. Mantas et al. (Eds.)*

*doi:10.3233/SHTI240649*

# Synergies Among Health Data Projects with Cancer Use Cases Based on Health Standards

Amelie GYRARD[a], Philip GRIBBON[b,1], Rada HUSSEIN[c], Somayeh ABEDIAN[c], Luis Marti BONMATI[d], Gibi Luisa CABORNERO[d], George MANIAS[e], Gabriel DANCIU[f], Stefano DALMIANI[g], Serge AUTEXIER[h], Rick van NULAND[i], Mario JENDROSSEK[j], Ioannis AVRAMIDIS[k] and Eva Garcia ALVAREZ[l]

[a] *Trialog, Paris, France*
[b] *Fraunhofer, ITMP, Hamburg, Germany*
[c] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
[d] *La Fe University Hospital Valencia, Valencia, Spain*
[e] *University of Piraeus Research Centre (UPRC), Athens, Greece*
[f] *Siemens, Brasov, Romania*
[g] *Monasterio Research Hospital (FTGM), Pisa, Italy*
[h] *DFKI, Bremen, Germany*
[i] *Lygature, Utrecht, Netherlands*
[j] *Health Data Hub, Paris, France*
[k] *Ubitech, Athens, Greece*
[l] *BBMRI ERIC, Graz, Austria*

ORCiD ID: Amelie GYRARD https://orcid.org/0000-0003-3938-5617

**Abstract.** We are creating a synergy among European Health Data Space projects (e.g., IDERHA, EUCAIM, ASCAPE, iHELP, Bigpicture, and HealthData@EU pilot project) via health standards usage thanks to the HSBOOSTER EU Project since they are involved or using standards, and/or designing health ontologies. We compare health-standardized models/ontologies/terminologies such as HL7 FHIR, DICOM, OMOP, ISO TC 215 Health Informatics, W3C DCAT, etc. used in those projects.

**Keywords.** Cancer, Health Data Space, Interoperability, Standards

## 1. Introduction

Cancer is a serious illness. Artificial Intelligence (AI) with Big Data and machine learning technological advances can help improve patients' quality of life.

We are creating a synergy among European Health Data Space (EHDS) projects (e.g., IDERHA, EUCAIM, ASCAPE, iHELP, Bigpicture, and HealthData@EU pilot project) via health standards usage thanks to the HSBOOSTER EU Project since they are involved or using standards, and/or designing health ontologies. We compare health-

[1] Corresponding Author: Philip GRIBBON; E-mail: philip.gribbon@itmp.fraunhofer.de.

standardized models/ontologies/terminologies such as HL7 FHIR, DICOM, OMOP, ISO TC 215 Health Informatics, W3C DCAT, etc. used in those projects.

## 2. Synergies among Health Data Space Projects

**IDERHA project** (HORIZON 2023-2028) facilitates heterogeneous health data access and reuse in **lung cancer** use cases. The IDERHA infrastructure connects existing data providers' resources so users can access and share the data, to analyze the data compliant with data standards, GDPR, and regulatory requirements. Existing (retrospective) datasets are reused for the development of analysis algorithms, and to evaluate and demonstrate the value and utility of the IDERHA platform to patients, researchers, regulators, and other stakeholders. Remote patient monitoring and data access consent management are embedded within mobile applications. IDERHA includes several categories of low and high-dimensional data: clinical and molecular data, image data, and patient-generated data types, from wearables measurements, PROMs and PREMs, and questionnaires. IDERHA offers solutions for data providers to share their citizen-controlled data, with other users for other needs. With the citizen-controlled data-sharing application, citizen users (whether patients or not) can share Personal Health Data with other parties for specific needs.

**iHELP project** (H2020 2021-2024) provides a personalized healthcare framework for **pancreatic cancer**, using AI and Big Data techniques for early risk prediction decision support and personalized prevention & intervention measures. It collects, integrates, and manages standardized data from heterogeneous health-related data sources (called Holistic Health Records (HHR)) [1][2], based on data modeling, management, and interoperability, and supports the continuous monitoring and analysis cycle. The HHR structure adheres to health informatics standards (e.g., HL7 FHIR) ensuring data is interoperable and accessible across different healthcare platforms and systems. Healthcare recommendations through mobile and wearable technology comply with data security standards (e.g., ISO/IEC 27001) and user privacy (e.g., GDPR). User-centric web, mobile, and wearable applications enable different healthcare stakeholders to utilize AI analytics and recommender system techniques – combined with simulation tools – to aid the delivery of personalized decision support and develop new healthcare pathways. iHELP validates its integrated platform in five different pilots 1) study of genomics and epigenomics markers for early risk assessment of pancreatic cancer, 2) interventional monocentric study based on patient-reported outcomes and assessing the toxicity risk of patients with pancreatic, prostate, and anal cancer, 3) the impact of lifestyle choices and tobacco on elevating the risk factors for pancreatic cancer, 4) risk, personalized recommendations, and measures to raise awareness of relevant factors for pancreatic cancer, and 5) risk prediction models and targeted interventions that can delay the onset of liver and pancreatic cancer.

**EUCAIM project (EUropean Federation for CAncer IMages)** is a pan-European digital federated infrastructure of cancer images (H2020 2020-2024), to develop AI tools for Precision Medicine. The infrastructure provides the validation and development of AI tools that support the cancer diagnosis procedure, treatment, and predictive medicine benefiting European patients. The infrastructure provides a platform for developing AI tools that aim to enhance cancer treatment and diagnosis procedures for patients. EUCAIM is aligned with the EHDS initiative to share a high-quality data and tool repository, including a hub hosting the Atlas of Cancer Images and a federated node that

preserves the data sovereignty of providers. EUCAIM assists clinicians, researchers, and innovators willing to support reproducible clinical decision-making systems for diagnosis, treatment, and predictive medicine. EUCAIM infrastructure enhances healthcare procedures for citizens and stimulates the European market by providing innovative tools and services. EUCAIM stores or provides access to anonymized data, including cancer images, and to a network of clinical sites with clinical information such as the type of tumor, molecular profile, treatment, and follow-up. It contributes to European policies and strategies to promote the secondary use of health data, focusing on patients with cancer. It will take essential steps towards the EHDS as it aligns with Europe's Beating Cancer Plan.

**ASCAPE project** (H2020 2020-2023), provides **HL7 FHIR profile related to oncology;** aims to take advantage of Explainable Artificial Intelligence, Federated Deep Learning, and Privacy-preserving data processing techniques to improve **cancer patients' quality of life** and post-treatment course. ASCAPE built an open AI infrastructure for health stakeholders (hospitals, health care administrators/providers, research institutions, etc.) to either deploy and run locally the AI algorithms on their private data, without sending them to the cloud or sending their data homomorphically encrypted to the cloud where they may be processed without the cloud decrypting them. Even though patient data remain private, as they are not shared with external parties, new knowledge emerges from several AI analytics shared through the open AI infrastructure. The ASCAPE project provides data-driven assistance to preserve and improve the Quality of Life (QoL) of cancer patients, datasets from cancer patients were used to train predictive models for QoL indicators. These models were used for predicting the quality of life evolution for specific patients and proposed interventions that benefit it. Within the context of ASCAPE project, two types of cancer were considered: **breast cancer and prostate cancer**, considering genders, and age groups. Data were collected from EHS systems of Research Hospitals in Spain, Greece, Sweden, and the UK, leveraging the usage of **HL7 FHIR standards** applied in a pivot conversion from the information model locally adopted in the healthcare provision, virtually defining an FHIR Implementation Guide for cancer patients [3]. ASCAPE designed and deployed intelligent intervention services for physiological and psychological support, and to improve patient and family counseling and guidance, early diagnosis and forecasts of ill health, identification of disease trajectories and relapse, and improved health literacy.

**Bigpicture** (H2020 2021-2027) **- Kidney:** It improves the quality and efficiency of toxicity studies and boost drug development for the benefit of patients. Bigpicture collects and shares quality-controlled pathology images, metadata, and AI algorithms. Building on existing assets derived from ELIXIR infrastructure, Bipicture established the first European ethical and regulatory-compliant platform connecting pathologists, AI developers, researchers, patients, and industry. Bigpicture is a catalyst in the digital transformation of pathology, allowing AI to reach maturity. It creates the tools and workflows to support the collection and sharing of images and metadata by European pathology departments. Bigpicture specified a data model that defines the Mandatory Submission Metadata (MSM) for directly Accessible Data Sets (MSMdad) and a Common Mandatory Metadata Structure (CMMS). The metadata storage format is based on the European Genome-Phenome Archive (EGA) format, however, it was expanded for the Bigpicture's needs which established a standardized value specification for all metadata attributes defined in the CMMS, based on surveys on the different encoding schemes and format specifications used by the different nodes and alignment to the EGA model and international published standards (e.g. SNOMED CT). In cases where no

standardized coding schemes exist, beneficiaries work on developing ontologies in the nodes. This is exemplified by non-neoplastic kidney disease, for which currently there is no unified coding available. Representatives of the renal node thus work with a group of neuropathologists on implementing Kidney Biopsy codes (KIBICO) into SNOMED-CT. Based on these results, a strategy is developed that allows a minimal set of ontologies, nomenclatures, and controlled vocabularies to be used in Bigpicture, while at the same time limiting the need for manual recording. To enable interoperability, it developed a technical implementation of file formats (DICOM) including the required tooling for data conversion to this common format. Bigpicture reports digital pathology guidelines, standards, and regulatory requirements. This work has contributed to an **ISO New Work Item Proposal for a standard on "Medical laboratories — Part 2: Digital pathology and artificial intelligence (AI)-based image analysis".**

**HealthData@EU pilot project (2022 - 2024) - (EHDS2) EU project** focuses on accessing and using health data challenges. There are five cross-border use cases: 1) infectious disease surveillance, Antimicrobial Resistance (AMR), 2) Thrombosis in COVID-19 patients, 3) COVID-19 testing, vaccination, and hospitalization, 4) cardiometabolic diseases, and 5) **colorectal cancer**. HealthData@EU addresses questions of data interoperability and quality to implement data discovery (set-up of a European metadata catalog, based on the Health DCAT-AP standard) and data access by developing a single data access application form as well as data preparation and data use.

## 3. Mapping Standards (HL7 FHIR, DICOM, OMOP, ISO, etc.) to Projects

We focus on a set of specific standards:

- HL7 FHIR, ISO/AWI TR 24305 Guidelines for implementation of HL7/FHIR based on ISO 13940 and ISO 13606.
- ISO 12052 DICOM (Digital imaging and communications in medicine).
- OMOP-CDM (agreement rather than a standard).
- Other Standards from ISO TC 215 and CEN TC 251 Health Informatics: ISO 13606 EHR Communication, ISO 27269 International Patient Summary, ISO/TR 12300 Principles of mapping between terminological systems, ISO 29585 Framework for healthcare and related data reporting, Data model compliant with ISO 12967-2:2020 HISA (Health Informatics Service Architecture).
- Other standards relevant to Knowledge Graphs, ontologies, etc.
    - W3C DCAT-AP
    - Standards for Semantic Web: W3C RDF, RDFS, OWL, SPARQL, etc.
- Other Standards such as ETSI SmartM2M SAREF for eHealth Ageing Well (https://saref.etsi.org/saref4ehaw/v1.1.1/)

We are even more interested when health standards produce health ontologies and with the online available ontology (Example: http://hl7.org/fhir/stu3/rdf.html). Table 1 provides a detailed view of standards usage in the projects.

**Synergies among Alliances and Other Projects:** Some synergies are being done with impactful alliances working on those topics, such as AIOTI Health WG. We are also involved in data space alliances such as BVDA, GAIA-X, etc. **AIOTI Health WG** released a White Paper IoT/Edge Computing and Health Data and Data Spaces Release 1.0 AIOTI WG Health March 2024, where IDERHA and HealthData@EU pilot projects

are cited among use cases. **Big Data Value Association (BDVA)** published the European Health Data Space white paper in July 2023.

**Table 1.** Study on health standards used in Health Data Space Projects

| **Projects related to cancer/ Standards** | **HL7 FHIR** | **DICOM** | **OMOP** | **ISO / CEN** | **Other Standards** | **Ontology Standards** |
|---|---|---|---|---|---|---|
| IDERHA/ Lung Cancer | HL7 FHIR | DICOM | OMOP | ISO TC 215 (Planned) | | DCAT-AP: HealthDCAT -AP (Planned) |
| Bigpicture Kidney | - | In use for all Whole Slide Images in the repository | - | Medical laboratories — Part 2: Digital pathology and artificial intelligence (AI)-based image analysis | - | - |
| EUCAIM Cancer Images | - | DICOM | OMOP | Image preparation, processing, data harmonization, segmentation and AI model predictions | - | - |
| iHELP Pancreatic | HHR based on FHIR | - | OMOP | Mapper transformation for data harmonization | ISO 27799:2016 | SNOMED, LOINC |
| ASCAPE Breast/Prostate | HL7 FHIR | - | - | ISO/CEN 13606 | - | LOINC, SNOMED |
| HealthData @EU Colorectal | Does not work on actually implementing data standardization based on common guidelines but rather observes and collects standardization efforts undertaken by research teams to help them in their research/work. DCAT-AP? Importance of FHIR Profiles. | | | | | |

## 4. Conclusion

This synergy among EHDS projects, thanks to the HSBOOSTER EU Project, focuses on the health standards usage including semantic interoperability by studying the usage of health ontologies. Tools are needed to support standards and help users choose the one fitting their needs to reduce the time-consuming task of comparing health standards, nomenclatures, data formats, etc. such as HL7 FHIR, DICOM, OMOP, etc. Future work will investigate the usage of health ontologies, and devices used to produce health data.

## References

[1] Manias G, Azqueta-Alzúaz A, Dalianis A, Griffiths J, Kalogerini M, Kostopoulou K, Kouremenou E, Kranas P, Kyriazakos S, Lekka D, Melillo F. Advanced Data Processing of Pancreatic Cancer Data Integrating Ontologies and Machine Learning Techniques to Create Holistic Health Records. Sensors. 2024 Mar 7;24(6):1739.

[2] Manias G, Kouremenou E, Alzúaz AA, Kranas P, Melillo F, Kyriazis D. An Optimized Pipeline for the Processing of Healthcare Data towards the Creation of Holistic Health Records. In2023 International Conference on Applied Mathematics & Computer Science (ICAMCS) 2023 Aug 8 (pp. 50-56). IEEE.

[3] Frid S, Fuentes-Expósito MA, Grau-Corral I, Amat-Fernandez C, Muñoz-Mateu M, Pastor-Duran X. Successful Integration of EN/ISO 13606–Standardized Extracts From a Patient Mobile App Into an Electronic Health Record: Description of a Methodology. JMIR Med Inform. 2022; 10 (10): e40344.