Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records

D3.1 – Data Modelling and Integrated Health Records: Design and open specification I

WP3 Personalised Holistic Health Records

Dissemination Level: Public
Document type: Report
Version: 1.0
Date: August 30, 2021

The project iHelp has received funding from the European Union's Horizon 2020 Programme for research, technological development, and demonstration under grant agreement no 101017441.

Document Details

Project Number	101017441
Project Title	iHelp - Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records
Title of deliverable	Data Modelling and Integrated Health Records: Design and open specification I
Work package	WP3
Due Date	31/08/21
Submission Date	30/08/21
Start Date of Project	January 1, 2021
Duration of project	36 months
Main Responsible Partner	Athens Technology Center (ATC)
Deliverable nature	Public
Author name(s)	Kalogerini Maritini (ATC), Athanasios Dalianis (ATC), Claudia Pandolfo (ENG), Fabio Melillo (ENG), George Manias (UPRC)
Reviewer name(s)	Pavlos Kranas (LXS), Ainhoa Azqueta (UPM)

Document Revision History

Version History			
Version	**Date**	**Author(s)**	**Changes made**
0.1	2021-02-16	Irida Manika (iSPRINT), Harm op den Akker (iSPRINT)	iHELP Deliverable Template – Initial version.
0.2	2021-03-10	Maritini Kalogerini (ATC)	Table of contents, Initial structure and text
0.3	2021-05-06	Maritini Kalogerini (ATC), Athanasios Dalianis (ATC)	Main body
0.4	2021-07-15	Fabio Melillo (ENG), Claudia Pandolfo (ENG)	Input
0.5	2021-07-20	George Manias (UPRC)	Input
0.6	2021-07-27	Pavlos Kranas (LXS), Ainhoa Azqueta (UPM)	Internal Review and comments
0.9	2021-08-27	Kalogerini Maritini (ATC)	Final updates in all sections
1.0	2021-08-30	Dimosthenis Kyriazis (UPRC)	Quality check and final version for submission

Table of Contents

List of Figures

Executive summary

This report provides an overview of the data modelling approaches adopted in iHELP project. The main goal is to capture the necessary data for personalized healthcare scenarios and 'transform' them into ready-to-use parameters into the Holistic Health Records (HHRs). Based on the outcomes of the CrowdHEALTH project[1], HHRs are structured health records that include several types of information relevant to patients' health status, such as medical and clinical data, laboratory measurements as well as data related to patients' social life and physical activity.

The basis for the mapping procedure is the utilization of the Fast Healthcare Interoperability Resources (FHIR) specification, which is a standard for electronically exchanging healthcare information and data, taking advantage of existing logical and theoretical models. FHIR is a new specification based on industry tactics, capitalizing the successes and challenges gained through defining and implementing standards such as the HL7 v1&v2 and CDA. The description and expansion of the HHR model will be based on an ontological level in order to accomplish a high degree of semantic interoperability. Moreover, taking into account that iHELP is focused on different factors associated with Pancreatic Cancer and relevant care pathways, appropriate standards will be used. For this reason, the mapping procedure is based on mCode standard, which focuses on concepts and fields from the oncological domain.

The ultimate goal is to achieve sharing of specific aspects of HHR between different health platforms taking advantage of key characteristics of HHR that will allow this inter-exchange, such as their scalability and extensibility.

[1] https://www.crowdhealth.eu/

1 Introduction

It is known that cancer is among the main cause of death and more specifically Pancreatic Cancer is the seventh leading cause of cancer-related-deaths, worldwide. Taking into account its high mortality rate as well as the lack of clear symptoms in its early stages, early risk assessment and early detection can be the key to reduce mortality significantly. Over time, personalized approaches to Cancer care have the potential to improve the duration of Cancer stages or even overall survival rate. For a trustworthy early risk assessment and detection, the utilization of clinical and social related data is of outmost importance [WCS, 2013].

However, these types of data come in different formats and data models and are stored in various databases. In order to facilitate the procedure of risk assessment of Pancreatic Cancer and to enable the harmonisation of heterogeneous datasets through a standardised model, a well-defined data mapping and modelling procedure is required. This deliverable provides an overview of the data modelling approaches adopted in iHELP project, where the basis for the mapping procedure is the utilization of the Fast Healthcare Interoperability Resources (FHIR) specification, in order to achieve a concrete description and expansion of the HHR model. In order to realise a high degree of semantic interoperability, the work described here follows an ontological approach.

Clinical care necessitates healthcare professionals to have access to patient record information. This information is probably not gathered in any database, but it may be distributed across multiple sources, held in different electronic formats or paper documentation, and it may also be represented as mixtures of narrative, structured, unstructured, encoded and multimedia entries. For example, almost 40% of American citizens will be diagnosed with cancer during their lifetime, but only 3% of adult cancer patients contribute in clinical trials, which provide most of the high-quality data, which are very important for cancer research. Today, there are more than 1,500 electronic health records (EHR) with incompatible data, limiting the valuable information cancer researchers could pull these records from. The exchange of EHR information is complex not only because of the fact that much of clinical meaning is derived not from individual data values, but also because of the way in which they are linked together as compound clinical concepts, grouped under a same heading-conditions or associated with preceding healthcare events during the act of data entry or data extraction. Taking advantage of a common standard and language for oncology, data from millions of combinations and comparison of treatment parameters could be used to improve and personalize the treatment of each patient. Nowadays, there is an urgent need for digitalizing Healthcare records, as they are in the middle of a modification and transformation driven by the convergence of biological in combination with information revolutions and by 'pressing' social changes. This need is commonly recognised as an obstacle to the effective delivery of health services, by clinical professions, by health iservice organisations and by governments internationally. Moreover, aspects of certainty, severity and the absence of findings requisite to be capable of rigorous and unambiguous representation [HL7, Append].

1.1 Scope of the document

This document summarizes the work done in iHelp project in terms of the modelling approaches. It describes the necessary data for personalized healthcare scenarios as well as the procedure for transforming them into ready-to-use attributes into the Holistic Health Records (HHRs).

1.2 Structure of the document

The document is divided into 6 main sections structured as follows:

- Section 1, includes the introduction and the main description of the document
- Section 2, gives an overview of the Electronic Health Records, providing information about their basic structure, requirements and architecture as these are defined in various case studies so far.
- Section 3, presents some of the most important standards in the Healthcare domain
- Section 4, describes the Holistic Health Record model that will be used in the context of the project. It starts by presenting its main principles and continues with the description of its main attributes that will be used as a baseline and the methodology that will be followed from the data gathering until its final formulation.
- Section 5, provides an overview of the current status in terms of data analysis and mapping towards the definition of the HHR model.
- Finally, Section 6 concludes this report.

1.3 Relevance with other Work Packages

This deliverable falls into the scope of WP3. The main objective of this work package is to shape, survey and extend the existing HHRs, in such a way so as to shape and implement the concept of personalized and integrated HHRs. At this stage of the project, there is an intense collaboration with WP6 ''Clinical Validation and Pilot Studies'' that is focused on verifying the applicability of the clinical and technical outcomes through representative pilot scenarios and providing useful feedback about the iHELP concepts and technologies. More specifically, each pilot partner from WP6 provides an initial list of the attributes that will be used for the mapping and modelling procedure in WP3. Moreover, in the latter stages of the project, the outcomes of the mapping procedure will assist the work to be done to release the mechanisms for scalable and efficient big data management, as well as the early risk assessment and personalised recommendations in WPs 4 & 5, ''Knowledge Management and Modelling in the iHELP Platform'' and ''AI for Early Risk Assessment and Personalised Recommendations'', respectively.

1.4 Background

A person-centred electronic health record (EHR) is an anticipated key to the demand of digitalizing Healthcare records. EHR are in the middle of a modification and transformation driven by the convergence of biological in combination with information revolutions and by 'pressing' social changes. The main challenge here is to provide to clinicians with an integrated and related view of the complete health and health care history of each patient. Towards a more digitalized era, patient's electronic health records must

be available, discoverable, understandable, structured and standardized as well. In fact, the usage of an electronic health record or EHR system offers much better control over information security. In other words, healthcare data must be able to support automated clinical decision support systems and other machine-based processing, shifting the role of informatics from reactive to proactive healthcare. This need is now widely recognised to be a major obstacle to the safe and effective delivery of health services, by clinical professions, by health service organisations and by governments internationally. The primary role of an EHR is to be used for purposes of setting objectives and planning the patients' care, documenting the delivery of care and assessing its outcomes (See Coming digital challenges in healthcare).

2 Electronic Health Record (EHR) Model

2.1 EHRs: Definition, Structure & Content

According to the International Organization for Standardization (ISO) definition, the EHR is a repository of patient data in digital form, stored and exchanged securely, being accessible by multiple authorized users. Electronic patients' records are useful both for hospitals and for general practice and they include both unstructured free text and coded data. Electronic questionnaires created by the health care providers, can be answered by the patients in a frame of questions that include their medical history, indications of food or medication intakes, pain intensity etc. Moreover, separated or integrated patient ''nursing'' information system has been developed, in order to support nursing documentation. As the time passes by, the structure and the content of EHRs varies and includes time-oriented data (the data are presented in chronological order), problem oriented data (each problem is described according to the subjective information, objective information, assessments and plan (SOAP)) and source oriented data (the content of the record is arranged according to the method by which the information was obtained,) as well as the combination of the three types [KHK, 2008]. EHRs include both unstructured free text and coded data, while in most EHRs the terminologies used in the records, such as their classification, their vocabularies etc. are described. Regarding the medical data components recorded in the EHRs, these include among others: referral, symptoms, previous medical history, life style and habits, physical activity, treatments, medication etc. EHRs are broadly used from people who provide health services that are mainly organized in three categories, the primary, the secondary and the tertiary care. Primary care is health care provided in the community by the staff of a general practice, secondary care is medical attention provided by a specialist facility upon referral by a primary care physician, and tertiary care is provided by a team of specialists in a major hospital [Med, 2005]. In this sense, the EHRs are used by health care professionals, by the administrative staff and last but not least by the patients or their assistants, given the fact that most patients nowadays also require access to their own EHR to an extent that permits them to play an active role in their health management. Many different global classifications are being used for purposes of nursing documentation (see Table 3). In addition, various other national classifications were also used in medical information documentation, including the ''Operationenschlussel nach'' coding for procedures, the Swedish coding system, the problem list vocabulary, the controlled terminology medical entities dictionary for problems, medications and adverse reactions and the drug dictionary for coding medication.

Table 1: The international terminologies used in EHRs

Data component	International terminology
Diagnoses	International Classification of Diseases (ICD)
	Read codes
	International Classification of Primary Care (ICPC)
Procedures	Current Procedural Terminology (CPT)
Medication	Anatomical Therapeutic Chemical Classification Index (ATC)
Pathological findings	Systematized Nomenclature of Medicine (Snomed)
Nursing problems	North American Nursing Diagnoses (NANDA)
	International Classification of Nursing Practice (ICNP)
Nursing interventions	Iowa Nursing Intervention Classification

	ICNP
Nursing outcomes	Iowa Nursing Outcome Classification (NOC)

2.2 Requirements for Representing and Communicating EHRs

The entries of proper health records are made as formal contributions to an evolving and flexible framework and at any time a patient's health record provides the information basis while new findings are interpreted. The following list presents some of the user and enterprise requirements that illustrate the needs in primary, secondary and tertiary care across countries and professions. The requirements were derived after many analyses from the expert groups internationally, concluding that the essential information that must be accommodated within an HER architecture, needs to:

- Capture the original meaning intended by a set of record entries
- Provide a framework appropriate to the needs of professionals and enterprises to analyse and interpret EHRs on an individual or population basis
- Incorporate the essential legal constructs to support the communication of EHR entries between professionals working groups, whilst respecting the privacy terms of individual patients

The communication of EHR information is complex because much of the clinical meaning is derived mainly from the way the data are linked in every health organization in order to form clinical concepts, thus this combination may not be uniformed across all organizations. Additionally, aspects like legal accountability impose the need for extra requirements in the definition and management of EHRs.

Some of the clinical and medico-legal contexts that needs to be considered during the mapping to classes and attributes in an EHR architecture are:

- Compositional context, e.g. provide labels for each content value or maintaining the initial data structure
- Content Value Context, e.g. usage of natural language, data value ranges, etc.
- Clinical Interpretation Context, e.g. additional comments by author, data presence/absence etc.
- Ethical and Legal Context, e.g. access rights, version control, etc.
- Care Process Context, e.g. provide links to other entities like process status, protocols etc.

2.3 Archetypes & the EHR Architectural Approach

Apart from the EU Third, Fourth and Fifth Health Telematics Framework Programmes, considerable research has been conducted over the past years in order to develop architecture formalisms for capturing healthcare data in an efficient way. The architectural approach is available in order to represent an EHR and it includes the concept of an 'Archetype'. An Archetype is the official definition of prearranged groupings of the building block classes defined in the Reference Model for specific clinical domains or organisations. The aim of an archetype is to specify and effectively constrain a precise hierarchy of record component, classes and sub-classes. Archetypes may be used within EHR systems to govern the EHR data held within a repository. The definition or constrain of the names, the attributes and other relevant values, optionality and multiplicity at any point in the hierarchy, the data types and value ranges or other possible dependency constraints, are described there. Finally, archetypes define and express the rules by which useful clinical templates can be constructed from the Reference Model in consistent and interoperable ways. The model under which the instances are conformed is called an Archetype Model. On the one hand, this model is

stable and predefined, but on the other hand, via version control, individual archetype instances can be revised or complemented by others, as long as the clinical practice evolves [SBK, 2012].

In the case that a set of EHR systems share a common set of archetypes, individual EHRs or decision support systems can request specific parts of one or more EHRs, from one or more EHR systems or a central repository. EHR systems will map the original clinical data to a consistent hierarchy of record components within an EHR extract. The potential sources of knowledge for developing such archetype definitions include:

- the data models of existing clinical systems
- data-entry templates or the lay-out of computer screen forms used for data entry
- shared-care data sets, messages and reports used locally or nationally
- the structure of templates and forms used for the documentation
- any pre-agreed terms in terminology systems.

3 Standards in Healthcare Data

3.1 Important medical data standards

Complex technical artefacts can be assembled out of smaller components, when standardizing health care data, the following actions should be involved:

1) Definition of data elements: A clear determination of the data content to be collected and exchanged should be performed.
2) Data interchange formats: Formats for electronically encoding the data elements - including sequencing and error handling, should be standardized. Document architectures for structuring data elements, information models that define the relationships can be included.
3) Terminologies: The medical terms and concepts should be noted. Those terms are used to describe, classify, and code the data elements. Also, data expression languages and syntaxes should be collected in order to describe the relationships among the terms/concepts.
4) Knowledge Representation: Information about standard methods, clinical & decision support guidelines, for electronically representing medical literature should be present.

Table 2: Important medical data standards

Standards development organisation	Standard	Scope
Federative Committee on Anatomical Terminology (FCAT)	Terminologia Anatomica (TA)	Anatomy terms in English and Latin
Health Level Seven (HL7)	v2	Messaging protocol; several of the chapters of this standard cover clinical content
	V3 (RIM)	Information ontology; especially the “Clinical Statement” work aims to create reusable clinical data standards
	CDA Level 1–3	Information model for clinical documents (Embedding of terminology standards in level 2 and 3); especially the Continuity of Care Document (CCD) specifications and the Consolidated CDA (C-CDA) specifications add detail to standards for clinical documents
	FHIR	Information and Document model; several parts of the core specification deal with clinical content
Integrating the Healthcare Enterprise (IHE)	Several Integration profiles	Clinical workflows including references to clinical data standards to be used
International Organization for Standardization (ISO)	TS22220:2011	Identification of subjects of care
	21090:2011	Harmonized data types for information exchange
	13606	High-level description of clinical information models

Standards development organisation	Standard	Scope
	23940 (ContSys)	Health care processes for continuity of care
	14155	Clinical investigations
	IDMP	Medicinal products
National Electrical Manufacturers Association (NEMA)	DICOM	Medical imaging and related data
openEHR foundation	openEHR	Clinical information model specification
Regenstrief Institute	LOINC	Terminology for lab and other observables
	UCUM	Standardised representation of units of measure according to the SI units (ISO 80000)
PCHAlliance (Personal Connected Health Alliance)	Continua Design Guidelines	Collecting data from personal health devices
SNOMED International, formerly knowns as the International Health Terminology Standards Development Organisation	SNOMED CT	Terminology / Ontology for representing the electronic health record ("context model" = Information model for SNOMED CT)
World Health Organization (WHO)	ICD-10 / ICD-11	Disease classification
	ICF	Classification of functioning, disability and health
	ICHI	Health procedure classification
	NN	Generic names for pharmaceutical substances
	ATC	Drug ingredient classification
World Organization of Family Doctors (WONCA)	ICPC	Primary care classification

3.1.1 Health Level Seven (HL7)

3.1.1.1 Fast Healthcare Interoperability Resources (FHIR)

Fast Healthcare Interoperability Resources (FHIR) Specification is a standard for exchanging healthcare information and data in an electronically form leveraging existing logical and theoretical models in order to make available a consistent and rigorous mechanism for exchanging data. The main goal behind FHIR is to build a set of resources that are able to satisfy the majority of common use cases as it allows healthcare information, including clinical and administrative data, to be findable. On one hand, FHIR can be used as a stand-alone data exchange standard and on the other hand, it can be used collaboratively with existing broadly used standards. Although HL7 standards have been producing healthcare data exchange and information modelling standards for over 20 years, FHIR is a new specification based on industry tactics, capitalizing the successes and challenges gained through defining and implementing standards such as the HL7 v2, HL7 v3 and CDA.

FHIR is based on internet standards generally used by industries, not necessarily in the healthcare domain. In particular, these include the REST approach, which describes how individual packets of information,

known as Resources, can be shared easily. In that scope, the main building block in the FHIR standard is the exchangeable content, named Resource. Resources, via their discrete information unit with well-defined meaning, can be collected for each type of clinical information. The defined Structured Data Elements that are logical, are mapped to formal definitions and can be presented in multiple formats (JSON, XML, Turtle). All resources share a common set of characteristics, which contain a common approach for their definition and representation, a common set of metadata and a human readable part. According to HL7 FHIR Release 5, there are 141 resources, categorized by maturity, by security categories, by Committee or by Standard Status. The FHIR Resources are categorized in 5 main categories and 4 or 5 subcategories, as it is shown below (Figure 1) [HL7].

FOUNDATION				
CONFORMANCE	**TERMINOLOGY**	**SECURITY**	**DOCUMENTS**	**OTHER**
CapabilityStatement	CodeSystem	Provenance	Composition	Basic Parameters
CapabilityStatement	ValueSet	AuditEvent	DocumentManifest	Binary
StructureDefinition	ConceptMap	Consent	DocumentReference	Bundle
ImplementationGuide	NamingSystem		CatalogEntry	Linkage
SearchParameter	TerminologyCapabilities			MessageHeader
MessageDefinition				OperationOutcome
OperationDefinition				Subscription
CompartmentDefinition				Topic
StructureMap				
GraphDefinition				
ExampleScenario				

BASE				
INDIVIDUALS	**ENTITIES #1**	**ENTITIES #2**	**WORKFLOW**	**MANAGEMENT**
Patient	Organization	Substance	Task	Encounter
Practitioner	OrganizationAffiliation	BiologicallyDerivedProduct	Appointment	EpisodeOfCare
PractitionerRole	HealthcareService	Device	AppointmentResponse	Flag
RelatedPerson	Endpoint	DeviceMetric	Schedule	List
Person	Location		Slot	Library
Group			VerificationResult	

CLINICAL				
SUMMARY	**DIAGNOSIS**	**MEDICATIONS**	**CARE PROVISION**	**REQUEST & RESPONSE**
AllergyIntolerance	Observation	MedicationRequest	CarePlan	Communication
AdverseEvent	DocumentReference	MedicationAdministration	CareTeam	CommunicationRequest
Condition (Problem)	DiagnosticReport	MedicationDispense	Goal	DeviceRequest
Procedure	Specimen	MedicationUsage	ServiceRequest	DeviceUseStatement
FamilyMember	BodyStructure	Medication	NutritionOrder	GuidanceResponse
History	ImagingStudy	MedicationKnowledge	VisionPrescription	SupplyRequest
ClinicalImpression	QuestionnaireResponse	Immunization	RiskAssessment	SupplyDelivery
DetectedIssue	MolecularSequence	ImmunizationEvaluation	RequestGroup	
		ImmunizationRecommendation		

SPECIALIZED				
PUBLIC HEALTH & RESEARCH	**DEFINITIONAL ARTIFACTS**	**EVIDENCE – BASED MEDICINE**	**QUALITY REPORTING & TESTING**	**MEDICATION DEFINITION**
ResearchStudy	ActivityDefinition	Evidence	Measure	MedicinalProductDefinition
ResearchSubject	ConditionDefinition	EvidenceVariable	MeasureReport	RegulatedAuthorization
	DeviceDefinition		TestScript	Ingredient
	EventDefinition		TestReport	ManufacturedItemDefinition
	ObservationDefinition			PackagedProductDefinition
	PlanDefinition			AdministrableProductDefinition
	Questionnaire			ClinicalUseIssue
	SpecimenDefinition			SubstanceDefinition
				SubstancePolymer
				SubstanceReferenceInformation
				SubstanceDefinition

FINANCIAL			
SUPPORT	**BILLIING**	**PAYMENT**	**GENERAL**
Coverage	Claim	PaymentNotice	Account
CoverageEligibilityRequest	ClaimResponse	PaymentReconciliation	ChargeItem
CoverageEligibilityResponse	Invoice		ChargeItemDefinition
EnrollmentRequest			ExplanationOfBenefit
EnrollmentResponse			Contract
			InsurancePlan

Figure 1: Important medical data standards

Each resource has a clear Scope & Usage definition paragraph, that explains where these resources are commonly used, what parameters they include and how they are described (name/value pairs, metadata etc.), among others. Moreover, the boundaries and relationships of this resource and the relevant boundaries are described in details. The content of these FHIR resources can be represented in different formats such as XML, JSON and Turtle, while other formats are also allowed.

For example, the description for the resource ''Observation'' that falls in the sub-category ''Diagnostics'', from the broad 'Clinical' category, includes paragraphs about the Scope and Usage that state that *''Observations are a central element in healthcare, used to support diagnosis, monitor progress, determine baselines and patterns and even capture demographic characteristics''*, [2] among other information. Furthermore, there is a list for the elements that the Uses of the Observation resource can include. More specifically, it can include:

- *Vital signs such as body weight, blood pressure, and temperature*
- *Laboratory Data like blood glucose, or an estimated GFR*
- *Imaging results like bone density or fetal measurements*
- *Clinical Findings* such as abdominal tenderness*
- *Device measurements such as EKG data or Pulse Oximetry data*
- *Clinical assessment tools such as APGAR or a Glasgow Coma Score*
- *Personal characteristics: such as eye-color*
- *Social history like tobacco use, family support, or cognitive status*
- *Core characteristics like pregnancy status, or a death assertion.*

The description about the Boundaries and Relationships explains where and how this resource should be used on the contrary to similar resources, as well as it declares the exact 'intention' of its creation. The content of the FHIR resources can now be represented in UML diagrams, XML template, JSON template & Turtle template. Using the example of the resource ''Observation'', it is possible to find information organised according to the FHIR data model, and represented in one of these formats, as it is illustrated in the following images (Figures 2-5).

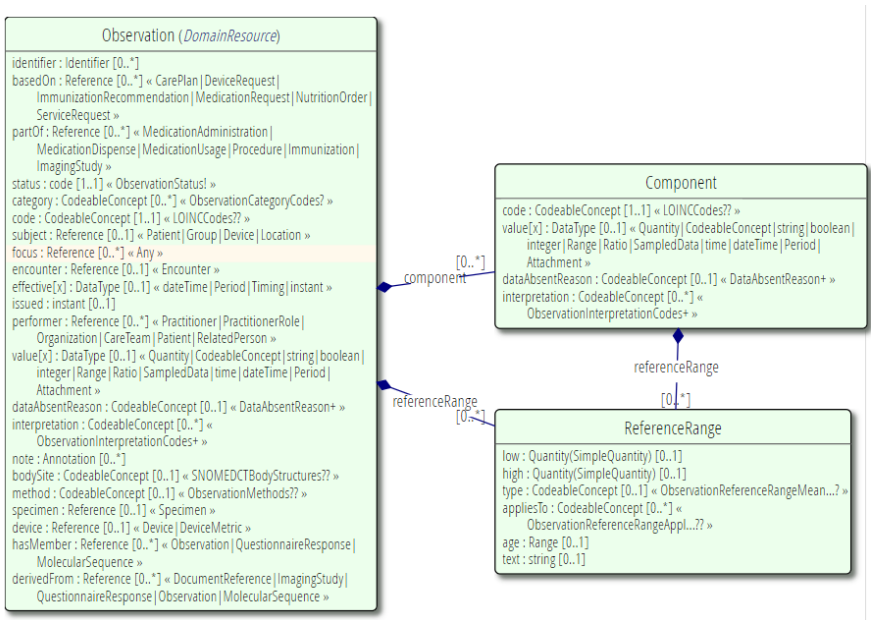

Figure 2: UML diagram of FHIR Resource ''Observation''

```
<Observation xmlns="http://hl7.org/fhir">
 <!-- from Resource: id, meta, implicitRules, and language -->
 <!-- from DomainResource: text, contained, extension, and modifierExtension -->
 <identifier><!-- 0..* Identifier Business Identifier for observation --></identifier>
 <basedOn><!-- 0..* Reference(CarePlan|DeviceRequest|ImmunizationRecommendation|
   MedicationRequest|NutritionOrder|ServiceRequest) Fulfills plan, proposal or order --></basedOn>
 <partOf><!-- 0..* Reference(ImagingStudy|Immunization|MedicationAdministration|
   MedicationDispense|MedicationUsage|Procedure) Part of referenced event --></partOf>
 <status value="[code]"/><!-- 1..1 registered | preliminary | final | amended + -->
 <category><!-- 0..* CodeableConcept Classification of  type of observation --></category>
 <code><!-- 1..1 CodeableConcept Type of observation (code / type) --></code>
 <subject><!-- 0..1 Reference(Device|Group|Location|Patient) Who and/or what the observation is about -->
</subject>
 <focus><!-- 0..* Reference(Any) What the observation is about, when it is not about the subject of recor
d --></focus>
 <encounter><!-- 0..1 Reference(Encounter) Healthcare event during which this observation is made --></en
counter>
 <effective[x]><!-- 0..1 dateTime|Period|Timing|instant Clinically relevant time/time-period for observat
ion --></effective[x]>
 <issued value="[instant]"/><!-- 0..1 Date/Time this version was made available -->
 <performer><!-- 0..* Reference(CareTeam|Organization|Patient|Practitioner|
   PractitionerRole|RelatedPerson) Who is responsible for the observation --></performer>
 <value[x]><!-- 🔒 0..1 Quantity|CodeableConcept|string|boolean|integer|Range|Ratio|
   SampledData|time|dateTime|Period|Attachment Actual result --></value[x]>
 <dataAbsentReason><!-- 🔒 0..1 CodeableConcept Why the result is missing --></dataAbsentReason>
 <interpretation><!-- 0..* CodeableConcept High, low, normal, etc. --></interpretation>
 <note><!-- 0..* Annotation Comments about the observation --></note>
 <bodySite><!-- 0..1 CodeableConcept Observed body part --></bodySite>
 <method><!-- 0..1 CodeableConcept How it was done --></method>
 <specimen><!-- 0..1 Reference(Specimen) Specimen used for this observation --></specimen>
 <device><!-- 0..1 Reference(Device|DeviceMetric) (Measurement) Device --></device>
 <referenceRange>  <!-- 0..* Provides guide for interpretation -->
  <low><!-- 🔒 0..1 Quantity(SimpleQuantity) Low Range, if relevant --></low>
  <high><!-- 🔒 0..1 Quantity(SimpleQuantity) High Range, if relevant --></high>
  <type><!-- 0..1 CodeableConcept Reference range qualifier --></type>
  <appliesTo><!-- 0..* CodeableConcept Reference range population --></appliesTo>
  <age><!-- 0..1 Range Applicable age range, if relevant --></age>
  <text value="[string]"/><!-- 0..1 Text based reference range in an observation -->
 </referenceRange>
 <hasMember><!-- 0..* Reference(MolecularSequence|Observation|
   QuestionnaireResponse) Related resource that belongs to the Observation group --></hasMember>
 <derivedFrom><!-- 0..* Reference(DocumentReference|ImagingStudy|
   MolecularSequence|Observation|QuestionnaireResponse) Related measurements the observation is made from
--></derivedFrom>
 <component>  <!-- 0..* Component results -->
  <code><!-- 1..1 CodeableConcept Type of component observation (code / type) --></code>
  <value[x]><!-- 0..1 Quantity|CodeableConcept|string|boolean|integer|Range|
    Ratio|SampledData|time|dateTime|Period|Attachment Actual component result --></value[x]>
  <dataAbsentReason><!-- 🔒 0..1 CodeableConcept Why the component result is missing --></dataAbsentReas
on>
  <interpretation><!-- 0..* CodeableConcept High, low, normal, etc. --></interpretation>
  <referenceRange><!-- 0..* Content as for Observation.referenceRange Provides guide for interpretation o
f component result --></referenceRange>
 </component>
</Observation>
```

Figure 3: XML template of FHIR Resource "Observation"

```
{
  "resourceType" : "Observation",
  // from Resource: id, meta, implicitRules, and language
  // from DomainResource: text, contained, extension, and modifierExtension
  "identifier" : [{ Identifier }], // Business Identifier for observation
  "basedOn" : [{ Reference(CarePlan|DeviceRequest|ImmunizationRecommendation|
   MedicationRequest|NutritionOrder|ServiceRequest) }], // Fulfills plan, proposal or order
  "partOf" : [{ Reference(ImagingStudy|Immunization|MedicationAdministration|
   MedicationDispense|MedicationUsage|Procedure) }], // Part of referenced event
  "status" : "<code>", // R!  registered | preliminary | final | amended +
  "category" : [{ CodeableConcept }], // Classification of  type of observation
  "code" : { CodeableConcept }, // R!  Type of observation (code / type)
  "subject" : { Reference(Device|Group|Location|Patient) }, // Who and/or what the observation is about
  "focus" : [{ Reference(Any) }], // What the observation is about, when it is not about the subject of r
ecord
  "encounter" : { Reference(Encounter) }, // Healthcare event during which this observation is made
  // effective[x]: Clinically relevant time/time-period for observation. One of these 4:
  "effectiveDateTime" : "<dateTime>",
  "effectivePeriod" : { Period },
  "effectiveTiming" : { Timing },
  "effectiveInstant" : "<instant>",
  "issued" : "<instant>", // Date/Time this version was made available
  "performer" : [{ Reference(CareTeam|Organization|Patient|Practitioner|
   PractitionerRole|RelatedPerson) }], // Who is responsible for the observation
  // value[x]: Actual result. One of these 12:
  "valueQuantity" : { Quantity },
  "valueCodeableConcept" : { CodeableConcept },
  "valueString" : "<string>",
  "valueBoolean" : <boolean>,
  "valueInteger" : <integer>,
  "valueRange" : { Range },
  "valueRatio" : { Ratio },
  "valueSampledData" : { SampledData },
  "valueTime" : "<time>",
  "valueDateTime" : "<dateTime>",
  "valuePeriod" : { Period },
  "valueAttachment" : { Attachment },
  "dataAbsentReason" : { CodeableConcept }, // C? Why the result is missing
  "interpretation" : [{ CodeableConcept }], // High, low, normal, etc.
  "note" : [{ Annotation }], // Comments about the observation
  "bodySite" : { CodeableConcept }, // Observed body part
  "method" : { CodeableConcept }, // How it was done
```

```
  "specimen" : { Reference(Specimen) }, // Specimen used for this observation
  "device" : { Reference(Device|DeviceMetric) }, // (Measurement) Device
  "referenceRange" : [{ // Provides guide for interpretation
    "low" : { Quantity(SimpleQuantity) }, // C? Low Range, if relevant
    "high" : { Quantity(SimpleQuantity) }, // C? High Range, if relevant
    "type" : { CodeableConcept }, // Reference range qualifier
    "appliesTo" : [{ CodeableConcept }], // Reference range population
    "age" : { Range }, // Applicable age range, if relevant
    "text" : "<string>" // Text based reference range in an observation
  }],
  "hasMember" : [{ Reference(MolecularSequence|Observation|
   QuestionnaireResponse) }], // Related resource that belongs to the Observation group
  "derivedFrom" : [{ Reference(DocumentReference|ImagingStudy|
   MolecularSequence|Observation|QuestionnaireResponse) }], // Related measurements the observation is ma
de from
  "component" : [{ // Component results
    "code" : { CodeableConcept }, // R!  Type of component observation (code / type)
    // value[x]: Actual component result. One of these 12:
    "valueQuantity" : { Quantity },
    "valueCodeableConcept" : { CodeableConcept },
    "valueString" : "<string>",
    "valueBoolean" : <boolean>,
    "valueInteger" : <integer>,
    "valueRange" : { Range },
    "valueRatio" : { Ratio },
    "valueSampledData" : { SampledData },
    "valueTime" : "<time>",
    "valueDateTime" : "<dateTime>",
    "valuePeriod" : { Period },
    "valueAttachment" : { Attachment },
    "dataAbsentReason" : { CodeableConcept }, // C? Why the component result is missing
    "interpretation" : [{ CodeableConcept }], // High, low, normal, etc.
    "referenceRange" : [{ Content as for Observation.referenceRange }] // Provides guide for interpretati
on of component result
  }]
}
```

Figure 4: JSON template of FHIR Resource ''Observation''

```
@prefix fhir: <http://hl7.org/fhir/> .


[ a fhir:Observation;
  fhir:nodeRole fhir:treeRoot; # if this is the parser root

  # from Resource: .id, .meta, .implicitRules, and .language
  # from DomainResource: .text, .contained, .extension, and .modifierExtension
  fhir:Observation.identifier [ Identifier ], ... ; # 0..* Business Identifier for observation
  fhir:Observation.basedOn [ Reference(CarePlan|DeviceRequest|ImmunizationRecommendation|MedicationReques
t|
  NutritionOrder|ServiceRequest) ], ... ; # 0..* Fulfills plan, proposal or order
  fhir:Observation.partOf [ Reference(ImagingStudy|Immunization|MedicationAdministration|MedicationDispen
se|
  MedicationUsage|Procedure) ], ... ; # 0..* Part of referenced event
  fhir:Observation.status [ code ]; # 1..1 registered | preliminary | final | amended +
  fhir:Observation.category [ CodeableConcept ], ... ; # 0..* Classification of  type of observation
  fhir:Observation.code [ CodeableConcept ]; # 1..1 Type of observation (code / type)
  fhir:Observation.subject [ Reference(Device|Group|Location|Patient) ]; # 0..1 Who and/or what the obser
vation is about
  fhir:Observation.focus [ Reference(Any) ], ... ; # 0..* What the observation is about, when it is not a
bout the subject of record
  fhir:Observation.encounter [ Reference(Encounter) ]; # 0..1 Healthcare event during which this observat
ion is made
  # Observation.effective[x] : 0..1 Clinically relevant time/time-period for observation. One of these 4
    fhir:Observation.effectiveDateTime [ dateTime ]
    fhir:Observation.effectivePeriod [ Period ]
    fhir:Observation.effectiveTiming [ Timing ]
    fhir:Observation.effectiveInstant [ instant ]
  fhir:Observation.issued [ instant ]; # 0..1 Date/Time this version was made available
  fhir:Observation.performer [ Reference(CareTeam|Organization|Patient|Practitioner|PractitionerRole|Rela
tedPerson) ], ... ; # 0..* Who is responsible for the observation
  # Observation.value[x] : 0..1 Actual result. One of these 12
    fhir:Observation.valueQuantity [ Quantity ]
    fhir:Observation.valueCodeableConcept [ CodeableConcept ]
    fhir:Observation.valueString [ string ]
    fhir:Observation.valueBoolean [ boolean ]
    fhir:Observation.valueInteger [ integer ]
    fhir:Observation.valueRange [ Range ]
    fhir:Observation.valueRatio [ Ratio ]
    fhir:Observation.valueSampledData [ SampledData ]
    fhir:Observation.valueTime [ time ]
    fhir:Observation.valueDateTime [ dateTime ]
    fhir:Observation.valuePeriod [ Period ]
    fhir:Observation.valueAttachment [ Attachment ]
```

```
  fhir:Observation.dataAbsentReason [ CodeableConcept ]; # 0..1 Why the result is missing
  fhir:Observation.interpretation [ CodeableConcept ], ... ; # 0..* High, low, normal, etc.
  fhir:Observation.note [ Annotation ], ... ; # 0..* Comments about the observation
  fhir:Observation.bodySite [ CodeableConcept ]; # 0..1 Observed body part
  fhir:Observation.method [ CodeableConcept ]; # 0..1 How it was done
  fhir:Observation.specimen [ Reference(Specimen) ]; # 0..1 Specimen used for this observation
  fhir:Observation.device [ Reference(Device|DeviceMetric) ]; # 0..1 (Measurement) Device
  fhir:Observation.referenceRange [ # 0..* Provides guide for interpretation
    fhir:Observation.referenceRange.low [ Quantity(SimpleQuantity) ]; # 0..1 Low Range, if relevant
    fhir:Observation.referenceRange.high [ Quantity(SimpleQuantity) ]; # 0..1 High Range, if relevant
    fhir:Observation.referenceRange.type [ CodeableConcept ]; # 0..1 Reference range qualifier
    fhir:Observation.referenceRange.appliesTo [ CodeableConcept ], ... ; # 0..* Reference range populatio
n
    fhir:Observation.referenceRange.age [ Range ]; # 0..1 Applicable age range, if relevant
    fhir:Observation.referenceRange.text [ string ]; # 0..1 Text based reference range in an observation
  ], ...;
  fhir:Observation.hasMember [ Reference(MolecularSequence|Observation|QuestionnaireResponse) ], ... ; #
0..* Related resource that belongs to the Observation group
  fhir:Observation.derivedFrom [ Reference(DocumentReference|ImagingStudy|MolecularSequence|Observation|
  QuestionnaireResponse) ], ... ; # 0..* Related measurements the observation is made from
  fhir:Observation.component [ # 0..* Component results
    fhir:Observation.component.code [ CodeableConcept ]; # 1..1 Type of component observation (code / typ
e)
    # Observation.component.value[x] : 0..1 Actual component result. One of these 12
      fhir:Observation.component.valueQuantity [ Quantity ]
      fhir:Observation.component.valueCodeableConcept [ CodeableConcept ]
      fhir:Observation.component.valueString [ string ]
      fhir:Observation.component.valueBoolean [ boolean ]
      fhir:Observation.component.valueInteger [ integer ]
      fhir:Observation.component.valueRange [ Range ]
      fhir:Observation.component.valueRatio [ Ratio ]
      fhir:Observation.component.valueSampledData [ SampledData ]
      fhir:Observation.component.valueTime [ time ]
      fhir:Observation.component.valueDateTime [ dateTime ]
      fhir:Observation.component.valuePeriod [ Period ]
      fhir:Observation.component.valueAttachment [ Attachment ]
    fhir:Observation.component.dataAbsentReason [ CodeableConcept ]; # 0..1 Why the component result is m
issing
    fhir:Observation.component.interpretation [ CodeableConcept ], ... ; # 0..* High, low, normal, etc.
    fhir:Observation.component.referenceRange [ See Observation.referenceRange ], ... ; # 0..* Provides g
uide for interpretation of component result
  ], ...;
]
```

Figure 5: Turtle template of FHIR Resource ''Observation''

3.1.1.2 HL7 RIM

The HL7 Reference Information Model is created by the standard organization Health Level Seven as a part of an internal release. This organization published the new version and as part of this release a reference data model is included, in order to serve as representation to collect and store specific clinical or administrative data, such as patient diagnoses, methods and costs for treatments as well as information concerning the personnel of a health organization. The main classes that compose the model include the classes: Act, Participation, Role and Entity. More specifically, a detailed description of these classes is listed below.

Act: This class includes instances regarding a specific action such as clinical or administrative of the sanitary environment at any given time. The status of these components can be characterized as planned, pending, completed, etc, and it can be of different types such as procedures, observations, drug administration's etc. as well as it involves different entities such as patients, health personnel and more.

Participation: The type as well as the degree of participation of different entities with different roles that may be involved in a clinical action, is related to the instances of this class.

Role: In this class, the functions of an entity that participates in a given action are included. It is noticeable that the same entity participates with various functions in the same act.

Entity: Each instance in this class represents any being, ranging from a living subject such as a patient to a sample of a microscopic organism. Moreover, chemical substances or physical devices are also listed here. The class "Act" is the most complex and important as well, and it is further divided into subclasses. Each one of these subclasses has unique attributes in order to satisfy the needs of each of these subclasses. The subclasses include the Observation, that indicates that the clinical action performed is an act of recognition, evaluation or indication of certain information about a subject, Procedure that indicates that the action

performed on the subject consists of some type of intervention or manipulation of part or parts of his body and substance Administration that Indicates actions to introduce or apply a particular substance or compound to a subject.

3.1.1.3 mCODE minimal Common Oncology Data Elements

Almost 40% of adults in America, will be diagnosed with cancer at least once during their lifetime, according to the National Cancer Institute. The ability to gain information from treatments of millions of patients, is supposed to be one of the most powerful tools against cancer. More specifically, if we had research-quality data from all cancer patients, it would enable higher quality health outcomes, but today there is an absence of data models, technologies, and methods to capture that information. Under that scope, mCODE™ (short for Minimal Common Oncology Data Elements) is a common standard and language for oncology, intended to assemble a core set of structured data elements for oncology electronic health records (EHRs). This standard would empower the treatment of every cancer patient to contribute to comparative effectiveness analysis (CEA) of cancer treatments and will allow effortless data exchange between health systems. mCODE has been created by the American Society of Clinical Oncology (ASCO®) in collaboration with the MITRE Corporation. mCODE is meant to be applicable to all types of cancer, but the initial emphasis has been on solid tumours. The clinical data requirements were drafted by a committee of twenty clinical experts in the fields of oncology, radiology, surgery, and public health. Under the umbrella of their knowledge, two use cases were developed with aim to drive the initial clinical data requirements for mCODE [mcode].

Use Case 1: Comparative Effectiveness Analysis and Cooperative Decision Making

Use Case 2: Comparative Effectiveness Analysis with Next Generation Sequencing (NGS)

Bellow, there is a list of several pre-existing standards that have been utilized in parallel with the information obtained from subject matter experts:

Table 3: List of Standards

List of pre-existing Standards, used to build mCODE
American Joint Committee on Cancer (AJCC) Staging Manual (8th Edition) Breast Cancer Chapter
College of American Pathologists (CAP) Cancer Protocols
North American Association of Central Cancer Registries (NAACCR) 2018 Site-Specific Data Items Manual
HL7 CDA R2 IG: Reporting to Public Health Cancer Registries from Ambulatory Healthcare Providers
National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology
RECIST Guidelines v1.1
HL7 FHIR Genomics Reporting Implementation Guidance (STU1)
The National Center for Biotechnology Genetic Test Registry (GTR)
The Human Genome Variation Society nomenclature for defining genetic variants.
The Human Genome Organization (HUGO) Gene Nomenclature Committee
The CDISC Therapeutic Area User Guides (TAUG)

The data elements identified were modelled using FHIR Shorthand (FSH) and SUSHI and exported as FHIR Profiles. At present, there are two distinct mCODE roles, the first role is the "mCODE Data Sender" and the second one is the "mCODE Data Receiver''. The first one provides data in response to a data query or autonomously pushes mCODE data to an mCODE receiver and the other accepts mCODE data from an mCODE Data Sender, accordingly. Regarding the Scope and Conceptual Model, this standard contains data elements divided into six groups:

a) Patient Group
b) Disease Characterization Group
c) Laboratory Results and Vital Signs Group
d) Treatments Group
e) Genomics Group
f) Outcomes Group

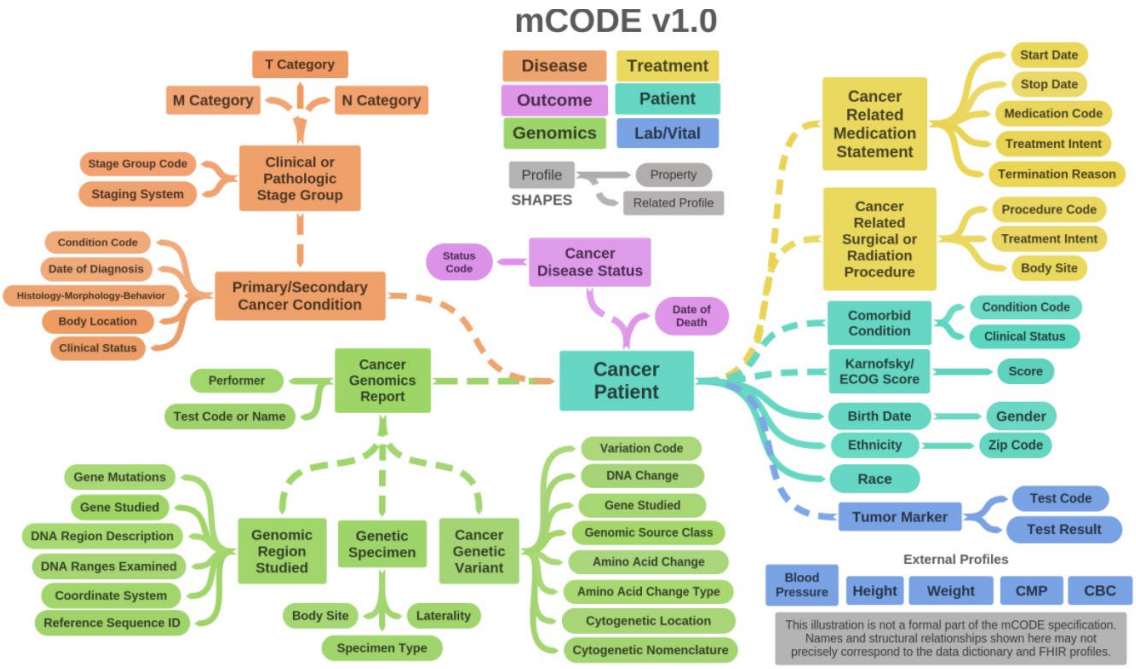

Figure 6: Data elements and groups for mCODE standard

a) The **Patient group** contains basic information about the patient, such as Demographics (including date of birth, gender, zip code, race, and ethnicity), Comorbid conditions ?, Patient performance status - It describes a patient's level of functioning in terms of their ability to care for himself, daily activity, and physical ability (walking, working, etc.), based on the Eastern Cooperative Oncology Group (ECOG) Performance Status and/or Karnofsky Performance Status (KPS) (Figure 7) [ECOG]. As performance assessments may be performed many times over a period of time, numerous instances may exist for a distinct patient.

ECOG PERFORMANCE STATUS	KARNOFSKY PERFORMANCE STATUS
0—Fully active, able to carry on all pre-disease performance without restriction	100—Normal, no complaints; no evidence of disease 90—Able to carry on normal activity; minor signs or symptoms of disease
1—Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work	80—Normal activity with effort, some signs or symptoms of disease 70—Cares for self but unable to carry on normal activity or to do active work
2—Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours	60—Requires occasional assistance but is able to care for most of personal needs 50—Requires considerable assistance and frequent medical care
3—Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours	40—Disabled; requires special care and assistance 30—Severely disabled; hospitalization is indicated although death not imminent
4—Completely disabled; cannot carry on any selfcare; totally confined to bed or chair	20—Very ill; hospitalization and active supportive care necessary 10—Moribund
5—Dead	0—Dead

Figure 7: The ECOG Performance Status and the Karnofsky Performance Status

b) The **Disease Characterization Group** contains data elements specific to the diagnosis and more specifically, to the staging of cancer. It includes the cancer diagnosis and more specifically the date and location of the cancer diagnosis, the tumor characteristics - the shape or behaviour of the tumor cell, the Cancer stage which describes the severity of a patient's cancer based on the magnitude of the primary tumor as well as on the extent of cancer metastasis. The cancer diagnosis combines three basic attributes: the type, the site, and certain characteristics of the cancer. mCODE supports three different code systems, as the use of these coding systems vary in different institutions, these are

-the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)

-the International Classification of Diseases, 10th version, Clinical Modifications (ICD-10-CM)

-and the International Classification of Diseases for Oncology, 3rd version (ICD-O-3)

The set of profiles contains information about the stage of the cancer, representing clinical stage group and pathologic stage group panels with members representing the primary tumor (T) category, the regional nodes (N) category, and the distant metastases (M) category. TNM staging systems are specified in the CancerStagingSystemVS extensible value set of SNOMED CT terms. Here, a single patient can have more than one staging panel, not common in practice.

c) The **Laboratory Results and Vital Signs Group** contains information from results deriving from the Complete Blood Count (CBC) (Automatic or Manual Differential) and the Comprehensive Metabolic Panel (CMP), that can be reported as individual laboratory observations or as grouped panels, using the DiagnosticReport resource. In this field, there are also the tumor markers attributes, which are key

prognostic factors in calculating cancer staging, identifying treatment options, and monitoring progression of disease. Moreover, vital signs are included- measurements of the most essential, or "vital" body functions such as blood pressure, heart rate, respiratory rate, and temperature, as well as height and weight. mCODE uses the FHIR vital sign profiles.

d) The **Treatments Group** includes reporting of actions and medications used to treat or try to treat a cancer patient. These treatments are derived from profiles such as:

-The CancerRelatedSurgicalProcedure, which represents surgical procedures that involve the removal of cancer tumors from the body.

-The CancerRelatedRadiationProcedure which is used to document the procedure of high-energy radiation from x-rays, gamma rays, neutrons, protons, and other sources to all cancer cells and shrink tumors.

- The MedicationStatement which records treatments involving chemotherapy agents, targeted therapy agents, and hormone therapy agents. In the mCODE profile of MedicationStatement there are also two extensions that distinguish it from FHIR's base resource of the same name (the TreatmentIntent - to record the purpose of the treatment & the TerminationReason - to document the reason for unplanned or premature termination of the medication). As far as the medication procedure is concerned, mCODE is fondness to medications used by the National Library of Medicine (NLM) RxNorm terminology, however it is limited to FDA-approved drugs and does not take account of drugs in the phase of clinical trials. To overcome this obstacle, mCODE permits for the inclusion of other coding systems like the NCI Thesaurus (NCIT).

e) Standardization of codes for genetic tests is vital to ease data analysis of genetic tests and should be a priority. Based on the HL7 CGWG Clinical Genomics Reporting Implementation Guide, the **Genomics Group in** mCODE includes a set of genomics related elements relevant to cancer assessment and treatment options. Even if the identity of non-genomic laboratory tests is typically represented by a Logical Observation Identifiers and Names (LOINC) code, many genetic tests and panels do not have LOINC codes. There are three profiles related to the capture of clinical genomics data:

- Genomics Report that include results of genomic analyses of different complexity and content, as simple as the results for a single discrete variant to complex sequences that are found in exome and genome-wide association studies (GWAS)
- Genetic Variant - to record variants that could be found from tests that broadly analyze genetic regions and stores results for any variants that could have been found.
- Genetic Specimen which assists in the further specification of the specimen collected for a genomics test.

f) The **Outcomes Group** is used to record outcomes of cancer treatment in mCODE and it involves two data elements: i) the disease status and ii) the date of death. The disease status is an analysis/assessment created by the oncologist that combines all available single type or multiple kinds of evidence about the patient, such as imaging data, assessment of symptoms, tumor markers, laboratory data, etc. According to mCODE, disease status is defined as "A clinician's qualitative judgment on the current trend of the cancer, e.g., whether it is stable, worsening (progressing), or improving (responding)''. The official recording of disease grade is every so often partial to clinical trials, involving precise criteria such as RECIST. The absence

of data outside of trials, restricts the application of everyday data, as disease status information is hardly found in organised forms of EHRs. On the other hand, the 'date of death' data can be found in many sources outside of the clinical setting. If this information is available in the EHR, it can be reported via mCODE, after the last clinical interaction. Many important relevant parameters can be derived from time-indexed observations of the status, such as: progression-free survival, time to recurrence, and overall survival. Finally, mCODE does not include patient reported outcomes such as the date of diagnosis.

Limitations of Minimal Common Oncology Data Elements

Several terminologies, including ICD-O-3 and the American Joint Commission on Cancer Staging System, as well as not cancer-specific such as Current Procedural Terminology (CPT) extensively are used in the cancer-related procedures such as surgeries or radiation. One restriction of mCODE is that it does not include content from these terminologies due to licensing restrictions. Moreover, mCODE elements may vary from the list identified by ASCO in their recent survey and these elements are under investigation and development based on reviewers from the oncology community. It should be noted that under Clinical Laboratory Improvement Amendments (CLIA) regulations, laboratory tests must include information on the performing technologist, performing laboratory, and performing laboratory medical director. While the performing laboratory can be determined by its resource type, in the current design of FHIR, there is no indicator that would discriminate the roles of the two Practitioner participants. Regarding the tumor marker tests, dissimilar to other laboratory profiles in mCODE, one profile has been created to handle them because of the large number of laboratory tests involved. A set of 150 tumor marker tests was established using an extensible binding to account for new tests and code updates. Finally, not all vocabularies used in mCODE are at this time supported by the FHIR Implementation Guide Publishing Tool.

3.1.2 International Organization for Standardization (ISO)

ISO is an International Organization for Standardization, which is currently a network of the national standards institutes of 157 countries[3]. They work on the basis of one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. ISO produces EHR standards that are limited to the structure and function of the EHR and the system that processes EHR. The primary purpose of ISO's family of EHR standards is to maximize interoperability between electronic records and systems that are specifically intended to be shareable, irrespective of the technologies they use and the platforms on which they reside. However, a variety of health information systems may include features and functionality that could be characterized as belonging to an EHR system. They help to ensure wide interoperability across global regulatory and healthcare communities, which is critical in ensuring accurate analysis and unambiguous communication across jurisdictions [SSC, 2019]. ISO has already published 37 standards relating to different aspects in the field of medical informatics across 22 countries [BEG, 2007]. The following seven sub-sections will introduce seven standards, which are concerned with the EHR interoperability as well as with the usage and utilization of data elements in the healthcare domain.

[3] https://www.investopedia.com/terms/i/international-organization-for-standardization-iso.asp

3.1.2.1 The TS22220:2011 Standard

ISO/TS 22220:2011 Health Informatics - Identification of Subjects of Health Care indicates the data elements and structure suited to accurate and procedurally appropriate and sensitive identification of individuals in health care in a face-to-face setting supported by computer technology, or through interactions between computer systems. It provides guidelines for improving the positive identification of subjects of care within and between health care organizations [ISO].

ISO/TS 22220:2011 defines demographic and other identifying data elements suited to capturing subject of care identification in health care settings, and the wide variety of manual and computer enhanced procedures used for this process. It provides guidance on the application of these procedures in the manual and the computer environment and makes recommendations about the nature and form of health care identifiers, the management organization to oversee subject of care identification and computer support to be provided for the identification process.

The objective of this standard is to promote uniform good practice in:

- Identifying individuals in a face-to-face, or paper-based environment, as well as in and between automated systems
- Recording and reporting of subject of care identifying data
- Ensuring that data being associated with any given subject of care, and upon which clinical communication and data aggregation are based, are appropriately associated with that individual or organization and no other

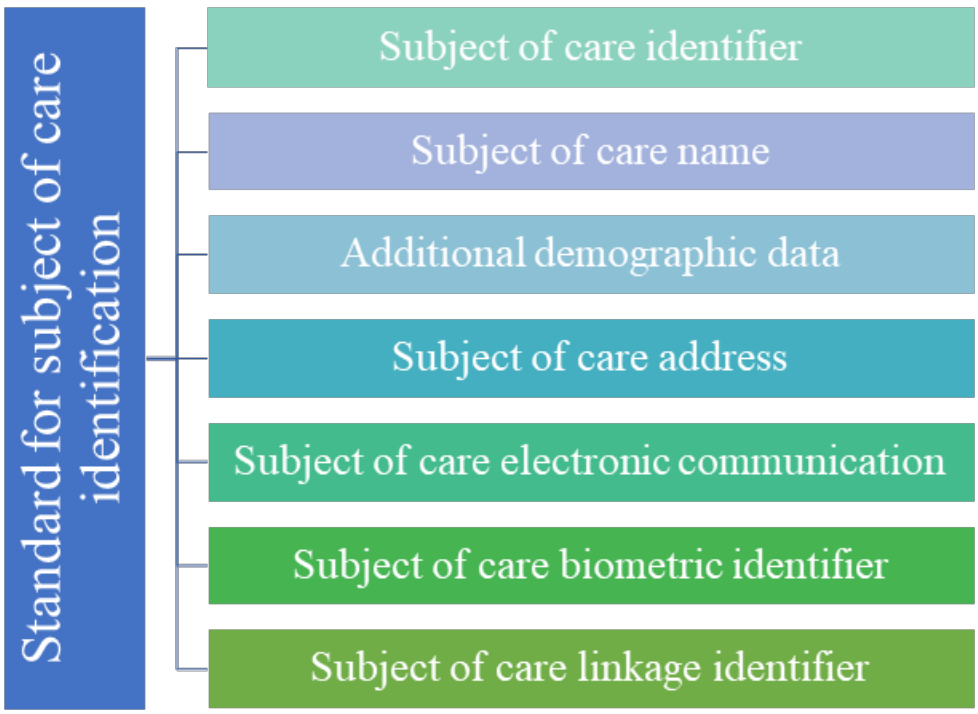

Figure 8: Data elements and interrelated components

3.1.2.2 The 21090:2011 Standard

ISO 21090:2011 Health informatics — Harmonized data types for information interchange Assistance from the Infrastructure and Messaging Committee in HL7 and the support of Connecting for Health have been instrumental in the preparation of this International Standard, which is a shared document between Health

Level Seven (HL7) and ISO, and has been produced according the terms of the agreement between HL7, CEN and ISO (JIC, see http://www.global-e-health-standards.org/), which ensures that the content is fully available through ISO, CEN and HL7 publication channels [CBS, 2015].

The ISO 21090:2011 standard also provides guidelines on applications of the model and makes recommendations about the nature and form of health care identifiers. More specifically it seeks to:

- provide a set of datatype definitions for representing and exchanging basic concepts that are commonly encountered in healthcare environments in support of information exchange in the healthcare environment;
- specify a collection of healthcare-related datatypes suitable for use in a number of health-related information environments;
- declare the semantics of these datatypes using the terminology, notations and datatypes defined in ISO/IEC 11404, thus extending the set of datatypes defined in that standard;
- provide UML definitions of the same datatypes using the terminology, notation and types defined in Unified Modelling Language (UML) version 2.0;
- specify an XML (Extensible Mark-up Language) based representation of the datatypes.

The requirements which underpin the scope reflect a mix of requirements gathered primarily from HL7 Version 3 and ISO/IEC 11404, and also from CEN/TS 14796, ISO 13606 (all parts) and past ISO work on healthcare datatypes. This International Standard can offer a practical and useful contribution to the internal design of health information systems but is primarily intended to be used when defining external interfaces or messages to support communication between them.

3.1.2.3 The 13606 Standard

ISO 13606 is a standard from the International Standardization Organization (ISO), originally designed by the European Committee for Standardization (CEN). The overall objective of the ISO 13606 standard is to define a rigorous and stable information architecture for communicating part or all of the electronic health record (EHR) of a single subject of care (patient) between EHR systems, or between EHR systems and a centralized EHR data repository. It may also be used for EHR communication between an EHR system and clinical applications or middleware components (such as decision support components) that need to access EHR data, or as the representation of EHR data within a distributed (federated) record system.

ISO 13606 considers the EHR to be the persistent longitudinal and potentially multi-enterprise or multi-national record of health and care provision relating to a single subject of care (the patient), created and stored in one or more physical systems in order to inform the subject's future healthcare and to provide a medico-legal record of care that has been provided [CMF, 2009].

Another remarkable feature of ISO/EN 13606 is that this standard not only consists of five different parts under the general title Health informatics - Electronic health record communication, but also that the alignment it presents to other relevant standards [TME, 2011]:

- **Part 1: Reference Model**. This part defines basic generic components that support information and the relationships between those components. This specific part can be seen as a subset of the openEHR Reference Model and presents a partial alignment with HL7 Clinical Document Architecture (CDA) Release 2.0. It can also be mapped to relevant portions of EN12796, EN13940, and specific metadata of the IHE Cross Enterprise Document Sharing (XDS).
- **Part 2: Archetype Model**. The Archetype Model represents the semantics of the dual model approach. An archetype is used for modelling domain concepts (blood pressure, body weight, etc.),

constraining the Reference Model at runtime by defining the structure of the instance and/or limiting the value range of an attribute. On top of this, this part leverages the openEHR model and its requirements have been adopted with minor revision by HL7. Moreover, the below figure shows a simplified scheme of the Archetype Model, extracted from ISO/EN13606-2. Describing a well defined archetype is not a simple task. As depicted in below figure, the ISO/EN 13606 standard offers different mechanisms to enable this modelling, such as the *archetype_description*, the *ontology* and the *constraint_model*.

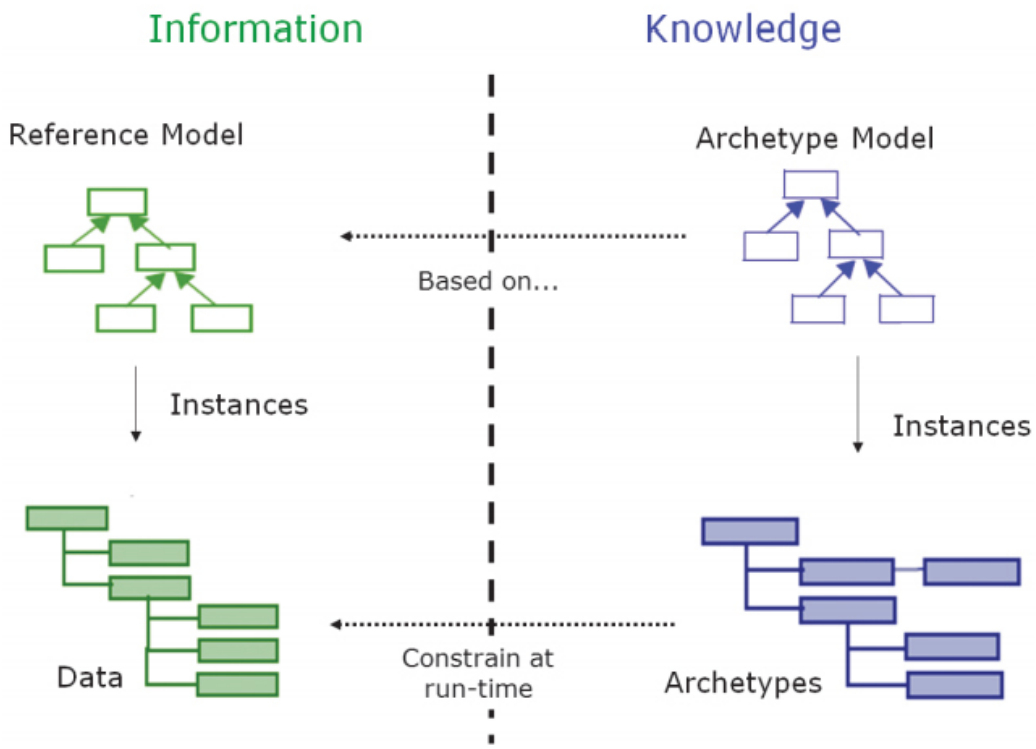

Figure 9: Relationship between information (instances of Reference Model) and knowledge (instances of Archetype Model)

- **Part 3: Reference Archetypes and Term lists**. This part establishes a normative set of coded terms, each one defining a controlled vocabulary for a Reference Model attribute contained in ISO/EN 13606-1. Also, this part contains mapping to HL7 Act Relationship codes, ACT_STATUS TERMS mapped with HISA, etc.
- **Part 4: Security**. This part describes a methodology for specifying the privileges necessary to access EHR data and some other general security requirements that should apply to HER communications. For example, it provides a double input table, the functional role of the requester and the sensitivity of the record. The information is only accessible if the functional role of the requester (coded with a number) is at least equal to the sensitivity of the record. This security part also defines both general and specific access policies able to deny or grant access to identified parties or specific functional roles. Moreover, Part 4 aligns with ISO 22600 and has been contributed to IHE in defining its privacy management services.
- **Part 5: Interface Specification**. This part describes a set of interfaces to request access to the information and resolve the request. Also, this part can be considered as a specialization of HISA services relating to clinical data and to clinical knowledge, and most parts of it can be mapped to IHE XDS query parameters.

Moreover, ISO 13606 defines a rigorous and stable information architecture for communicating part or all of the Electronic Health Record (EHR) of a single subject of care (patient). This is to support the interoperability of systems and components that need to communicate (access, transfer, add or modify) EHR data via electronic messages or as distributed objects; preserving the original clinical meaning intended by the author and reflecting the confidentiality of that data as intended by the author and patient. To achieve this objective, ISO 13606 follows a Dual Model architecture. On top of this, this Dual Model architecture is represented by Part 1 and Part 2 respectively, that have also been introduced latter in this section, and defines a clear separation between information and knowledge. The former is structured through a Reference Model that contains the basic entities for representing any information of the EHR. The latter is based on archetypes, which are formal definitions of clinical information models, such as discharge report, glucose measurement or family history, in the form of structured and constrained combinations of the entities of a Reference Model. The combination of the Reference Model (to represent data instances) and the Archetype Model (to semantically describe those data) provides a powerful capability of evolution to the information systems. Clinical information models (archetypes) may change in the future, but data will always remain interoperable. The internationally adopted strategy to overcome the interoperability gap is the application of standards that provide descriptions of all the elements involved, such as syntactic, structural and semantic interoperability. Thus, the Reference Model sets hierarchical relationships between its components, achieving in this way syntactic interoperability, i.e., identifying different elements in the system and establishing rules for combining them, thus allowing any system to be able to understand the structure of the information.

3.1.2.4 The 13940 (ContSys) Standard

The EN ISO 13940 System of Concepts for Continuity of Care is a single part standard that defines concepts around the topic of planned co-operation between various healthcare providers inside and outside of one jurisdiction, the patients and its surrounding. The clinical process and the concepts needed for all aspects, especially continuity, of the clinical processes are its focus.

The concepts in ContSys are defined solely from the enterprise/clinical perspective. The clinical context is represented in ContSys by a process view and a generic clinical process model. All concepts are directly or indirectly related to the clinical process model, which gives traceability to the clinical context. For the continuity of care 'clinical' is defined as any relation between subjects of care and healthcare professionals [TC,2013].

The aim for this standard is to provide a comprehensive, conceptual basis for content and context in healthcare services. It is the foundation for interoperability at all levels in healthcare organisations. All concepts are drawn in UML schemata to show the relations between the concepts.

This standard will not be directly implementable, but several steps of concretizing are needed before ContSys can be used in practice. One of the issues identified was the time required to understand the terms used in the standard, which are intentionally neutral and therefore not always easily recognised by those who use more common terms, albeit in a less rigorous way.

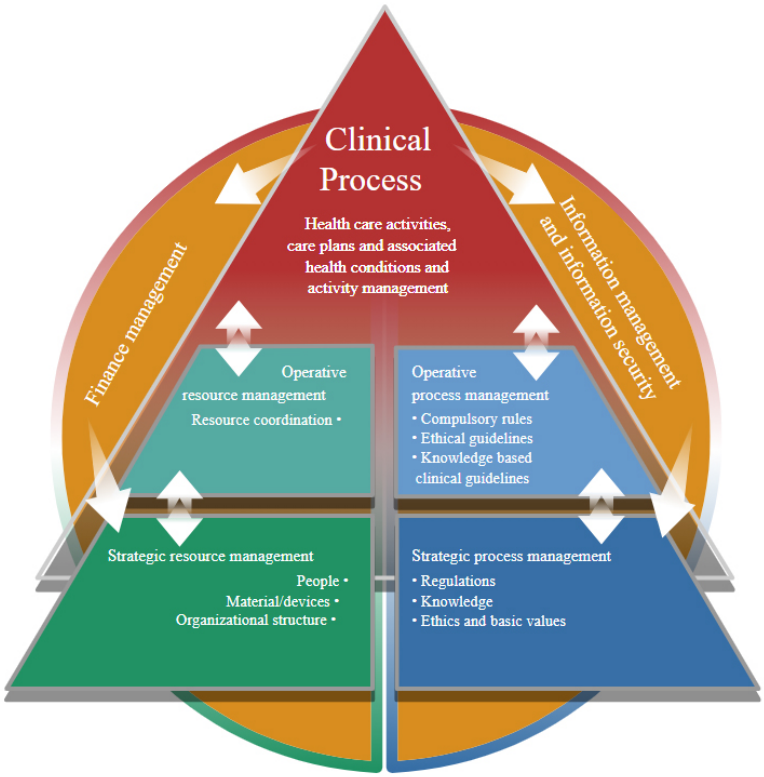

Figure 10: A visualisation of a model conforming to International Standard EN ISO 13940:2016

3.1.2.5 The 12967 (HISA) Standard

The HISA standard was developed by CEN Technical Committee (TC) 251, the technical committee for Health Informatics within the federation of European national standards bodies (CEN). The CEN HISA standard was adopted by the International Organization for Standardization (ISO) in 2009, with the stated aim of ISO 12967 being to provide guidance on:

- the description, planning and development of new electronic health systems; and
- the integration of existing electronic health systems, both intra- and inter-organizationally, through architecture that integrates common data and business logic into middleware, which is then made available throughout whole information systems.

The Health Informatics Service Architecture (HISA) is a 3-part standard (EN ISO 12967) which specifies a unified and open service architecture based upon a middleware of information services. These services are independent from specific applications/technology and should be capable of integrating common data and business logic services in any health information system (including EHR systems) [KSE, 2007].

The ISO 12967 series specifies fundamental requirements for 'information infrastructure' and provides guidance for the description, planning and development of new systems as well as for the integration of existing information systems, both within one enterprise and across different healthcare organizations through an architecture integrating the common data and business logic into a specific architectural layer (i.e. the healthcare specific service architecture), distinct from individual applications and accessible throughout the whole information system through information services, as shown in the below figure (Figure 11).

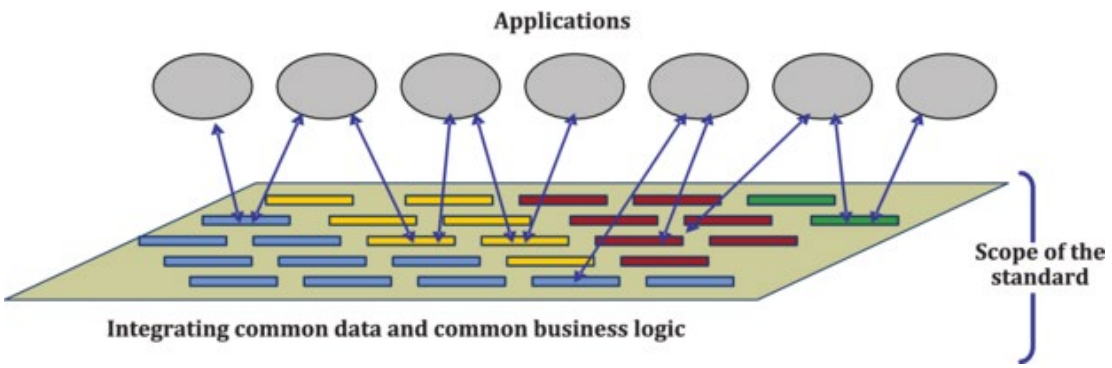

Figure 11: Scope of the ISO 12967 series

Moreover, EN/ISO 12967 is broken down into three parts and viewpoints: Enterprise Viewpoint; Information Viewpoint; and Computational Viewpoint, all of which deal with different aspects of ensuring service architecture supports openness and vendor-independence [ISO, 2009].

- **Part One - Enterprise Viewpoint:** The Enterprise Viewpoint component of EN/ISO 12967 specifies a set of fundamental common requirements at enterprise level with respect to the organizational purposes, scopes and policies that should be supported by the information and functionality of the middleware. In addition to this it provides direction for the integration of existing information systems, within the one enterprise and across different healthcare organizations. Part one of the standard sets forth the common enterprise-level requirements (e.g. workflows, authorizations) that must be supported through the HISA, which integrates the common data and business logic into a specific architectural layer (i.e. the middleware), accessible throughout the whole information system of the health service.
- **Part Two - Information Viewpoint:** The Information Viewpoint component of EN/ISO 12967 sets forth the fundamental characteristics of the information model to be implemented by the middleware to provide comprehensive, integrated storage of the common enterprise data and to support the fundamental business processes of the healthcare organisation, as defined in ISO 12967 Part One. The specifications were designed to be universally relevant, whilst being sufficiently specific to allow implementers to derive an efficient design of the system for their organisation. This specification does not aim to provide a fixed, complete specification of all possible data that may be necessary for any given health service. It specifies only a set of characteristics, in terms of overall organisation and individual information objects, identified as fundamental and common to all healthcare organizations.
- **Part Three - Computational Viewpoint:** The Computational Viewpoint component of EN/ISO 12967 specifies the scope and characteristics of the services that should be provided by the middleware for allowing access to the common data as well as the execution of the business logic supporting the enterprise processes identified in the information viewpoint. The computational model, like the information model is designed to be universally relevant, whilst still being sufficiently specific to allow implementers to derive an efficient design of the system for their organisation, irrespective of the specifics of the pre-existing information technology environment in which it will be implemented.

3.1.2.6 The 14155 Standard

ISO 14155 Clinical investigation of medical devices for human subjects -- Good clinical practice addresses good clinical practices for the design, conduct, recording and reporting of clinical investigations carried out in human subjects to assess the safety and performance of medical devices for regulatory purposes. However, it does not apply to in vitro diagnostic medical devices [ISO, 1996].

This standard was developed by ISO technical committee ISO/TC 194 and its first edition was published on 15 February 2003. ISO 14155 was published in its second edition in February 2011, while the third edition was released in July 2020.

The ISO 14155 standard provides the general specifications and requirements for clinical investigations to:

- Protect the rights, safety, and well-being of human subjects;
- Ensure scientific conduct of the clinical investigation and credibility of the clinical investigation results;
- Define the responsibilities of the sponsor and principal investigator;
- Assist sponsors, investigators, ethics committees, regulatory authorities, and other bodies involved in the conformity assessment of medical devices.

In any stage of the clinical development, from early feasibility studies and pre-market clinical investigations to post-market studies and registries, medical device manufacturers should comply with the ISO14155 standard to ensure ethical and scientific conduct of the investigation and credibility of the clinical study results [JKJ, 2018].

To this end, conducting a clinical investigation according to the ISO 14155:2020 standard provides a high level of protection for human research subjects and allows companies to perform international device trials that generate clinical data with high integrity that are accepted by regulatory agencies worldwide [KM, 2010].

3.1.2.7 The ISO IDMP Standards

IDMP is a suite of five standards developed within the International Organization for Standardization (ISO). These standards provide an internationally-accepted framework to uniquely identify and describe medicinal products with consistent documentation, coding and exchange of product information between global regulators, manufacturers, suppliers and distributors. The IDMP suite of standards are a result of a need to standardize the definition of medicinal product and substance information to facilitate the unique identification and exchange of such information in the context of pharmacovigilance. Initially, this methodology was created in the context of pharmacovigilance. Then it was extended to all other areas in healthcare, where information on medicines is needed, such as in prescribing and dispensing of medicinal products [GN, 2017]. To ensure broad synergies between global regulatory and medical communities, these standards have been developed and published under the auspices of ISO with the participation of the International Council for the Harmonization of Technical Requirements for Pharmaceutical Products for Human (ICH), Health Level Seven (HL7), as well as other international stakeholders and experts[4].

To this end, the ISO IDMP consists of five separate standards published in 2012 that describe data elements and their structural relationships for unique identification and exchange of information [KGF, 2020]:

- **Medicinal Product Identification (MPID) - ISO 11615:** Data elements and structures for unique identification and exchange of regulated medicinal product information. MPID describes the

[4] https://www.ema.europa.eu/en/human-regulatory/overview/data-medicines-iso-idmp-standards-overview

detailed data elements and their structural relationships required for the unique identification of regulated medicinal products. Data elements that identify and characterize a medicinal product include the product name (authorized by regulatory agency), clinical particulars (e.g. indications, contraindications), pharmaceutical product (substance, dosage form, route of administration), medicinal product packaging, marketing authorization (e.g., authorization number, application information), manufacturer/establishment, etc.

- **Pharmaceutical Product Identifier (PhPID) - ISO 11616:** Data elements and structures for unique identification and exchange of regulated pharmaceutical product information. PhPID uniquely associates medical products with the same or similar pharmaceutical composition based on the following data elements: substance(s), strength(s) (units of measurement/presentation), reference strength(s), and dosage form.
- **Substance Identification (SubID) - ISO 11238:** Data elements and structures for unique identification and exchange of regulated information on substances. Defines substances that constitute a medicinal product by their main, general characteristics.
- **Dosage Form and Route of Administration - ISO 11239:** Data elements and structures for unique identification and exchange of regulated information on pharmaceutical dose forms, units of presentation, routes of administration and packaging.
- **Units of Measurement (UoM) - ISO 11240:** Data elements and structures for unique identification and exchange of units of measurement. Specifies rules for the usage of units of measurement, establishes requirements to provide for traceability to international metrological standards, defines requirements for representation of units of measurement in coded form, provides structures and rules for mapping between different unit vocabularies and language translations.

The primary purpose for the development of these suite of standards was in response to a worldwide demand for internationally harmonized specifications for medicinal product identification. IDMP specifies data elements, structures, and relationships, in addition to required terms and definitions. There is a vast amount of information required in a dossier and IDMP is not intended to be the complete solution for all regulatory data represented. However, the ability to uniquely identify a medicinal product and structure this information per these standards allows for consistent implementation throughout the international regulatory framework—to achieve regulatory harmonization and ultimately convergence [PVA, 2017].

The benefits of IDMP suite include:

- **Safety Surveillance:** Unambiguous global identification will improve pharmacovigilance by uniquely identifying specific medicinal products in Individual Case Safety Reports (ICSRs). Globally detect safety signals from medicinal products referenced in adverse events.
- **Transparency:** Enhances the opportunity to communicate and build trust with the public and other stakeholders about medicinal product quality and safety.
- **Mitigation of Drug Shortages:** Standard facilitates the identification of pharmaceutically equivalent products across regions, to support mitigation of drug shortages.
- **Interoperability:** Harmonized source for product information based on vocabularies and standards that are consistent across the globe. Moreover, it seeks to support the exchange of medicinal product information between companies and regulators.

Moreover, ISO IDMP covers the entire medicinal product lifecycle, including products in development, investigational products, products under evaluation and authorised products. Thus, by using ISO IDMP within regulatory activities brings benefits to regulators, industry and, ultimately, to the patients.

3.2 The eHealth Standards Roadmap

If the development and full adoption of eHealth tools and solutions in healthcare delivery in Europe is described as a journey, it requires a roadmap. In 2015-2017, via the eStandards initiative, there was an attempt to develop a roadmap in order to foster the development as well as the adoption of eHealth standards and specifications. Driving by the vision of a global exchange of these data, the building of consensus on interoperability procedures across health related data standards is of major importance in Europe and beyond. In an evidence-based roadmap, these initiatives correlated with different clinical use cases for various paradigms and embedded a quality management system for interoperability testing and certification of eHealth systems. Specific demands- such as the increasing need, expectation, and cost of healthcare resulting from ageing populations, the high investment in new drugs, any changes in doctor-patient relationship- which requires better access to information about patents' health and the preferred options for care and treatment, increased demand for home-based and mobile care available- can be core components of a health services changes. However, for such eHealth directions more than data from local pilots is required, so that services can interoperate, be scaled-up and remain sustainable within a healthcare system. The practice for standards development and for the creation of a precise roadmap for adopting a specific set of standards could not be achieved without a continuous flow between three acts of design, development, and interaction: Co-creation, Governance and Alignment [KPD, 2019]. Co-creation includes:

- Co-design of services – co-planning of health and social policy, co-prioritisation of services and co-financing of services, co-commissioning;
- Co-delivery of services – co-managing and co-performing services
- Co-assessment – co-monitoring and co-evaluation of services.

Governance Standards are closely linked to the governance of healthcare systems and healthcare workflows and it is both a regulatory and a political process that involves balancing competing influences and demands. And finally, alignment, is the element that ensures that changes in the perceptions of stakeholders or changes in governance are accommodated into projects and initiatives already underway. It is anticipated that the application of the eStandards practise in an iterative way, with focus to reusable interoperability components, specification and tools, with dynamic governance, will advance global health data interoperability, which is the ultimate goal.

4 Holistic Health Record (HHR) model

The notion of Holistic Health Records (HHRs) has been initially realised during the CrowdHEALTH project [KAB, 2018].

HHRs is a conceptual model aiming at capturing clinical events and laboratory test results, but also properties about lifestyle, social care, personal measurements, environment, social interactions and in general all health determinants that can be relevant for health risk detection and personal healthcare. Such data may be collected or produced by medical operators, by citizens themselves or by automatic sensors.

The original HHR designed in CrowdHEALTH will be adapted to include relevant data for iHelp, such as those related to pancreatic cancer, and comply with the pilot needs.

4.1 Main Principles

The need to define Holistic Health Records has been driven by two main reasons:

- Health-related data is currently being produced, collected and stored in various formats across multiple types of systems. In addition to typical medical systems at hospitals, medical centres and health organisations, new ICT services (e.g., those using IoT devices, sensors and mobile applications for health monitoring or decision support) also collect vast amount of health-related data. Currently all these systems operate independently: data exists either as free-text or structured data, that may be organised in a proprietary way or be coded using one-of-many coding, classification or terminologies that have often evolved in isolation and designed to meet the needs of the context that they have been developed in. This has resulted in many data integration and interoperability issues making very difficult to provide a centred approach for health monitoring, risk predictions, preventions and personalised interventions.
- Despite the huge number of alternative standards currently existing in the health domain, e.g. DICOM, HL7 CDA, HL7 v2, HL7 v3, HL7 FHIR, openEHR, CEN/ISO EN13606, covering clinical information produced by health professionals and exploited by health institutions worldwide for their EHR (Electronic Health Record) systems, none of them completely supports the possibility to represent non-clinical events. Furthermore, these models do not allow to differentiate clinical information recorded by the patient from information recorded by the health organization and usually assume they are recorded by the latter.

The HHR model is built upon the following principles

1. It aims at consistently representing all data required by a specific use case
2. It should be extensible, by including data types currently not needed but likely to be in the future.
3. It is defined using the FHIR standard as the main reference. Although FHIR primarily represents clinical data, also nonclinical data, such as information collected from sensors could be described using this standard. Furthermore, using *resources* and flexible extension mechanisms, the FHIR model can be employed in different situations. Entity types related to specializations and abstractions of FHIR elements can be qualified using ontologies.
4. It is designed in a modelling language, and concurrently is mapped with existing standards. To ensure the viability of a direct mapping to FHIR there are several constraints for designers, further described below, but a direct reference to FHIR has not been used for two reasons:
    - to unbind the HHR model from some assumptions adopted by FHIR and

- to simplify the use of HHR model separately from FHIR by explicating some aspects currently implied in FHIR

As a result, the HHR model aims at being easily implementable both on top of existing FHIR implementations and using other technologies.

The principal HHR features, and constraints are:

- Overall, each class of the HHR model is related to a resource type or a data type of the FHIR model, yet the HHR model is more specialized than the FHIR model because the application context is more definite. Classes and elements existing in FHIR, but not required by the project's use cases, are missing in the HHR model, which instead includes additional attributes/associations related to extensions of the FHIR standard. Also, an HHR class corresponding to a specific FHIR resource class may have explicit subclasses that are not represented as distinct resource classes in FHIR, because in the HHR model these explicit subclasses do not need a related FHIR extension. The HHR model explicitly define concepts required by the project's use cases but implicit in FHIR or needing a FHIR extension.
- The higher specialization of the HHR model makes it less ambiguous and more understandable, thus minimizing the risk that different standard elements are used to represent the same type of information (this possibility is likely to happen with standards, such as FHIR, providing by design different alternatives to represent the same information).
- The HHR model is designed at an ontological level. This approach leads to two differences with the FHIR model:
    - the multiplicity constraints on the properties and attributes of the entities/classes represent real word existence constraints, and not integrity constraints, as stated in FHIR. For example, the value 1 for the minimum multiplicity of an attribute indicates that at least one value of that attribute always exists in the world, even if this information is not stored in any IT system or not transmitted (that should be done according to the FHIR model)
    - HHR uses abstract classes with no direct corresponding type in FHIR, but with related super-types of FHIR resource types. The introduction of this type of classes was driven by the need of make explicit some implicit FHIR semantic commonalities.
- The HHR model uses some specific stereotypes and patterns to indicate ontological distinctions that cannot be expressed in standard UML. For example, classes of entities (e.g. Patient) representing roles of instances of other classes (e.g. Person), are labelled with the stereotype <role> and use the standard relation "player" to associate the entity (. e.g. the person) that plays the role. In OWL annotations can be used instead of stereotypes.
- When a class *C* has several subclasses not introducing specific attributes or constraints, then the subclasses are reified[5]. Each subclass is represented by an item of an enumeration (using stereotype <enum> in UML and the owl:oneOf property in OWL) and a mandatory attribute of the class *C* (with name *C*type) is used to represent the specific subclass of the instance. For example, the subclasses of the class *Condition* correspond to values of the enumeration *ConditionType* and the specific subclass of a *Condition* instance is represented by the value of the attribute named *conditionType*.

[5]In semantic terminology, reification is the process of turning a predicate into an object.

- For the purpose of ensuring an easy mapping with FHIR and with specific coding systems, some constraints are enforced to the HHR model. The principal constraint is that any leaf element of the HHR model (i.e. any class, attribute or association that does not have subclasses or specializations) must correspond to exactly one (resource or data) type of the FHIR model, i.e. all possible instances of an HHR class must represent the same entities of possible instances of only one corresponding FHIR class. On the other side, to avoid using instances of non-leaf classes to represent unintended entities, any non-leaf element of the HHR model is considered ontologically "abstract", i.e. all its representable instances or values must be instances or values of some subclass. Another constraint is that each instance of a HHR class must correspond to exactly one instance of the FHIR model.
- In order to provide more evidence of the mapping with FHIR, the name of the most general HHR element mapped to a specific FHIR element is usually the same of the corresponding FHIR element. However, when the semantics of the HHR element is indeed too specific, using the same name would be confusing and different names are more suitable.

4.2 Personalized parameters

Most of the standards listed in the Section 3 "Standards in Healthcare Data" are exclusively related to health data, and more specifically to the health data collected and exchanged among health care institutions and stored in the so-called Electronic Health Records (EHR). Nowadays this approach can be extended, since the patients are more engaged in their own health and the spreading of domestic device useful for monitoring the health status. The individuals can enter their own health data in the system called Personal Health Record (PHR) that can collect those data as a repository, in the simplest idea. But the PHR followed in the years different implementations [PAD, 2006]. The simplest approach is a standalone system, not interconnected, that the individual can use for storing and retrieval health information.

More advanced approach, with a significant increase in complexity, is a remote PHR, shared with the health organisations' EHR. The EHR system, optionally, can incorporate the data shared by the patients using the PHR system. Usually, the application or web application used by the patients to collect data is provided by the same organisation that provides the EHR service in order to overcome the interoperability issue of data integration.

The data types that the PHR can collect could be conceptually resumed in this table:

Table 4: PHR conceptual data type categories

Data Type	Manual input	Automatic input
Problem list	x	
Procedures	x	
Major Illness	x	
Allergy data	x	
Self-monitored data	x	x – needed smart devices
Family history	x	
Lifestyle	x	x – needed smart devices
Medications	x	
Laboratory tests	x	x – needed interconnected laboratories

Immunisations	x	x – needed interconnected laboratories

The data can be collected manually and automatically, for instance using some smart devices or importing interoperable laboratory results. Other than that, some data type are defined as subjective if inputted by the patients, such as the symptoms, the descriptions of the medical problems, the response to the administrated questionnaires [CJJ, 2006].

In this context, there is the concrete possibility to include the PHR data inside the certified EHR system, differentiating the sources, but providing the clinicians with a more detailed big picture of the status of the patients, in terms of health, but also including lifestyle, self-monitored parameters etc.

That integration will be only possible if the EHR systems and the PHR systems will use a common data model that can allow the exchange and the aggregation of data among different structures and personal devices. The HHR seeks to reach that goal being independent from proprietary and customised protocols and standards, although being compliant with the standard HL7/FHIR data format; so that the interoperability with systems adopting such a standard is guaranteed by design.

More than this, in iHelp project the HHR model aims at modelling also the data coming from the social interactions, so that the AI provided by the platform can also use those set of data to look for new correlations among data that may point out potential impacts on the health status of individuals.

Finally, modelling patient provided data to be included in the platform, and adding also the mentioned new sources (e.g. social network interactions) could provide several benefits for different stakeholders:

- Patients: can access their own data on request, can be prompted or motivated from the system to reach a goal, have a test or simply answer a questionnaire.
- Clinicians: can read all the available data in a single place, can be assisted by the AI to estimate the risk of developing diseases and are always in control of the next actions to be taken.
- Model Builder: can elaborate more and more sophisticated algorithm to highlight and present risks to the clinicians, counting on an interoperable and extendable data model (HHR).

4.3 Methodology followed to define the HHR model

The incremental approach, followed in the iHelp project, also drives the development of the HHR model, which will be defined in several issues, producing versions aligned with the collected requirements and datasets available. Some phases should be carried out, for each internal development cycle.

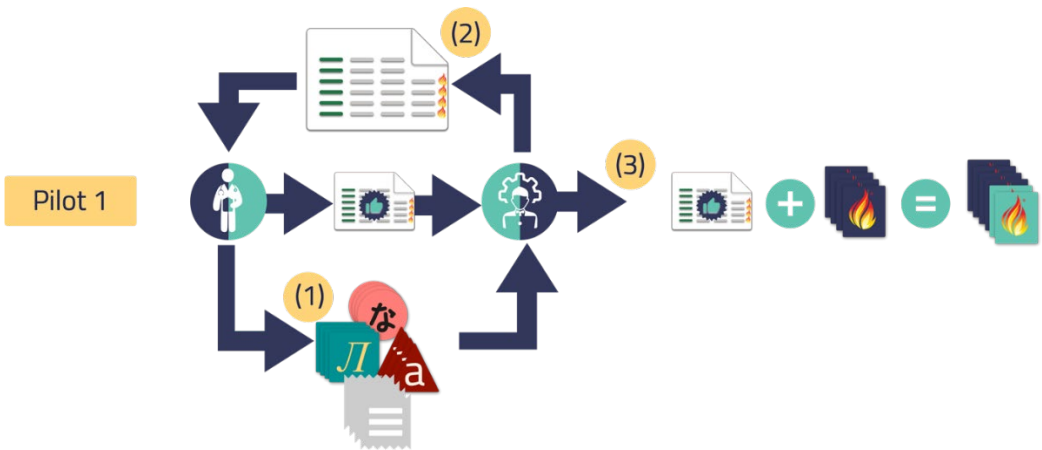

Figure 12: First steps, for each pilot

1. In principle, the technical partners with the role of data architect collect from the pilot reference partners the structure of the data that the organisations would like to store in the iHelp Big Data Platform and needed for the AI to be analysed, to compute the expected models. The expected description contains the data entities, their attributes, and their reciprocal relationships.
2. Those data description is then analysed by the data architects who produce a spreadsheet for expressing formal details. In this phase, the descriptions provided by different pilots are analysed and compared searching for similarities, inconsistencies, ambiguities, or lack of required information. If necessary, more details are requested to the pilot partners in this phase, such as the actual medical dictionary/terminology/ontology used to represent their data types, the cardinality of the attributes, constraints if any. To ease the interpretation of the provided data description, each dataset is individually mapped to FHIR v4 [HL7] and mCode standards. When a field or entity cannot be mapped to these standards, fields from relevant ontologies will be used and if this is not possible either, the field will be provided a new ontological definition in the HHR model.
3. The pre-processed data descriptions obtained during the previous step are then compared with the existing HHR resources and the coverage of the new datasets are verified together with the needs of extensions of the existing HHR model.
4. The mappings produced from the analysis of every pilot data description is then merged in a common HHR model. The different conceptual entities used by the pilots in different ways but resulted in FHIR resources with the same semantics, are mapped in a merged HHR conceptual class, or in a subclass of the same concept. A similar analysis is performed also on the attributes and relationships of those classes. This common HHR model is documented and shared using a notation like UML 2.0 [OMG,UML] or OWL.
5. Finally, the definition of the HHR model and the mapping with FHIR is formalised using a language like RDF/XML or JSON-LD.

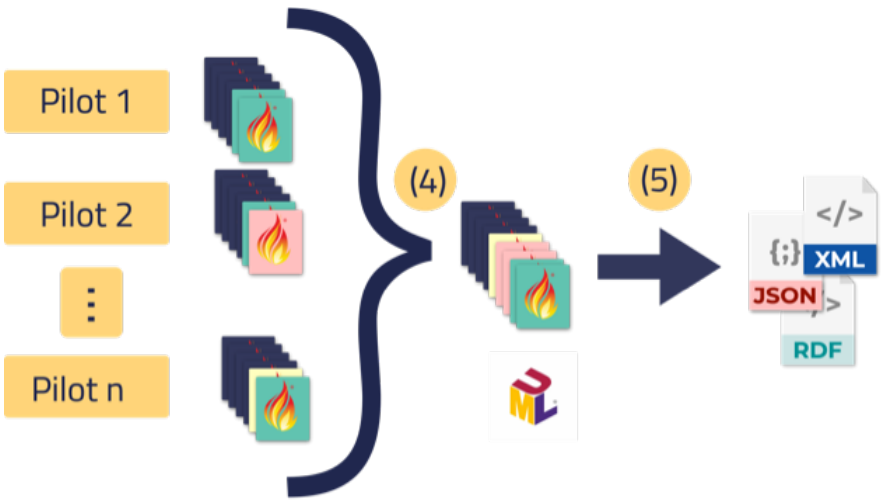

Figure 13: Mapping, merging and formalisation

4.4 HHR baseline

The HHR model contains data types representing all significant health determinants for an individual health status, that can be produced by different human actors or systems, in different moments of the patient's life, and considers both medical data, associated with regular patient care or as part of a clinical program,

and non-medical data that may have an impact on the patient's health status. It's called "holistic" because it is a data model not limited to clinical information, but that represents different health-related aspects of the citizens, overall, together providing a comprehensive view of the health conditions and associated risks for the patients.

The HHR data types will be organized in several categories like the following. The list of these categories will be finalised at a later stage of the project when the analysis of the user requirements and the provided data is also completed (see Figure 14 and Figure 15):

- **physical activities**: workouts, biodata and fitness tests performed by a person or groups of persons.
- **lifestyle**: data regarding sleep, daily habits (sedentary, active or vigorous) and substances consumption, e.g., alcohol, tobacco and so
- **social**: data concerning social interactions, namely emotions, contacts and exchanged multimedia items.
- **events:** aspects related to episode of care, hospitalizations, clinical procedures, laboratory tests and care plans.
- **medications:** all data regarding the prescription, request and assumption of medication.
- **conditions:** symptoms, diagnosis, allergies and intolerances suffered by a specific patient or group of patients.
- **nutrition:** all data pertaining to food and beverage intakes.
- **administrative:** demographics and other administrative information about an individual (patient, practitioner or student) or group of individuals, including also details on the educational level, occupational status and assurance of individuals, and information about organizations (school, municipality, region etc.);
- **measurements:** measurements and simple assertions about a patient, device or other subject, together with collective health measurements about a group of persons sharing common characteristics (e.g., living in the same town, district etc).

Figure 14: HHR covered health related aspects: Physical activities, lifestyle, social and events

Figure 15: HHR covered health related aspects: Medications, conditions, nutrition, administrative and measurements

The HHR model is a conceptual model, independent from specific implementation architectures. By preserving the distinction between conceptual and physical view, the HHR model aims at ensuring the integration of the data from heterogeneous sources and the possibility to implement it on top of existing FHIR libraries, being at the same time usable independently from FHIR (and its future evolutions) and applicable also for different ambitions: unlike FHIR, mainly focused on data exchange, the HHR model is intended to also support data integration and analysis.

The present work will be based on the outcomes of the CrowdHEALTH project regarding the definition of the HHR model, which provides the baseline for the work to be carried out in iHelp project. In the context of the iHelp project the HHR model will be altered, extended, and enhanced to not only cover the needs of the project's pilots but also to provide a more standardized version / suggestion for the medical community to use. Having this in mind, we need to also consider the following:

- In iHelp, the need for semantic representation of concepts, especially in the context of WP3, is greater than CrowdHEALTH. This means that describing the HHR model in an ontological language like OWL, would make more sense as it provides a higher degree of semantic interoperability.
- The advent of the mCode standard, allows us to describe concepts and fields from the oncological domain, thus concepts not defined in CrowdHEALTH or defined using custom ontologies can now be included in the HHR model in a more standardized way.
- Although the HHR model is independent from any implementation architecture, the aim is to learn from the lessons in CrowdHEALTH and provide a refined version that will also take into consideration the technical effort needed to realise and use the HHR model.

5 Data mapping

In this section we present the work done so far in defining the HHR model for the iHelp project. As mentioned in Section 4 there are several steps towards the completion of the HHR model. These steps may involve several repetitions, as clarifications and or additions of fields maybe necessary. At this point of the project, we gather the data definitions from the pilots and analyse the data in order to perform the necessary mappings.

5.1 Data Samples

The pilots fill in their data descriptions in an excel, providing the attribute names, their type, a description of these attributes, possible values, if they are mandatory or not as well as additional notes, grouped in categories.

The data provided by the pilots vary in size and nature of the data and their categories, e.g. the data provided by TMU so far are grouped in the following categories:

- Department Basic Information
- Physician Basic Information
- Patient Basic Information
- National Death Registry
- OPD Basic Information
- Registration Basic Information
- Inpatient Appointment & Record
- Inpatient Basic Information
- ER Record
- Medication Basic Information
- Patient Drug Allergy Record
- OPD Medication Record
- Inpatient Medication
- Inpatient Medical Order Log
- Inpatient Diagnosis Record
- Inpatient External Wounds Diagnosis Record
- Inpatient Disease Classification
- Inpatient Reservation & Diagnosis
- Examination Results
- Operation Basic Information
- Operation Diagnosis Record
- Cancer Registration
- Chemo Patient Record
- Chemo Medication Usage Record
- Chemo Side Effect Assessment
- Cancer Treatment Curative Effect
- Tumor Type
- Pathology Report

- Medical Supplies Basic Information

While the data provided by Gemeli for example are grouped in the following categories:

- Patient info
- Diagnosis info
- Tumor info
- Chemotherapy
- Survival

Some of the common categories found so far are:

- patient info,
- tumour info,
- diagnosis info,
- chemotherapy info,
- treatment / medication etc.

Indicatively, we provide in the following tables, fragments of these data descriptions.

Table 5: TMU - Patient basic information

Attributes	Type	Description	Values
CHR_NO	String	Patient ID	Encrypted
HOSP_TYPE	Integer	Hospital type	1: TMUH 2: Wanfang Hospital
PAT_NAME	String	Patient name	Encrypted
ID_NO	String	Patient national ID	Encrypted
PAT_ID	Integer	Patient identity	0: 1: Health insurance 2: Self payment
BIRTH_PLACE	String	Patient birthplace	code_type='05'
AC_TYPE		Patient birth year (in Taiwanese year)	
BIRTH_DATE	Date	Birthday	YYYMMDD (The format of all of the rest dates in this document are the same as this one.)
SEX_TYPE	String	Gender	1: Male 0: Female (The format of all of the rest gender in this

Attributes	Type	Description	Values
			document are the same as this one.)
BLOOD_TYPE		Blood type	
RH_TYPE		RH type	
EDU_CODE	String	Education status	code_type='04'
MER_FLAG	String	Marriage status	y:yes n: no
AREA_CODE1	String	Postal code	
HOME_TEL_NO	String	Home tel number	Encrypted
AREA_CODE2	String	Postal code	
OFFICE_TEL_NO		Office tel number	
WORK_CODE	String	Occupation	code_type='20'
ALLI_FLAG	String	Allergic to drugs	y:yes n: no
ZIP_CODE1	String	Zip code	code_type='07'
ADDR_NAME1	String	Address	Encrypted
ZIP_CODE2	String	Zip code	code_type='07'
ADDR_NAME2	String	Address	Encrypted
CONT_REL	String	Relations between patient and contact person	code_type='RT'
CONT_NAME	String	Emergency contact	Encrypted
CONT_AREA_CODE	String	Postal code	
CONT_TEL_NO	String	Contact person tel number	Encrypted
CONT_ZIP_CODE	String	Contact person zip code	code_type='07'
CONT_ADDR	String	Contact person address	Encrypted
DISC_CODE	String	Discount identity code	code_type='0L'
PART_CODE	String	Copayment code	code_type='23'
RTN_FLAG	String	Patient record archive	y:yes n: no

Attributes	Type	Description	Values
FST_OPD_DATE	Date	First visit date	
RECE_OPD_DATE	Date	Recent revisit date	
OWE_AMT	Double	Due amount	
DISP_CNT	Date	Skipping appointment time	
RECE_DISP_DATE	Date	Recent skipping appointment time	
LOC_CODE	String	Location of Medical record	
PRE_LOC_CODE	String	Latest location of medical record	
ACT_FLAG	String	Stagnancy Medical record	y:yes n: no
MORE_FLAG	String	Numerous medical record	y:yes n: no
DEATH_FLAG	String	Medical record of deceased patient	y:yes n: no
LEND_DATE	Date	Latest lend date	
LEND_TYPE	String	Latest lend type	1,3,4,6,7,8,9,B, (Blank)
LEND_NAME	String	Latest lend name	
DEPT_CODE	String	Latest lend department code	
SHIFT_NO	String	Latest lend time	(Blank)/1: morning/2: afternoon/3: evening
LOC_DESC	String	Latest lend location	
CONT_FLAG	String	Medical record in control	y:yes n: no
VIP_CODE	String	VIP code	type='33'
BACK_DATE	Date	Return date	
ID_CODE	String	Birth place code	
HOME_CODE	String	Address code	type='CB'
LIVE_CODE	String	Residence code	type='CB'

Attributes	Type	Description	Values
MERG_CHR	String	Merging to Medical record	
TRAN_CHR	String	Medical record before conversion	
ALLI_REASON	String	Allergy record	
SER_ILL	String	Serious illness record	
HEMO_DATE	Date	Latest health examination	
PEP_DATE	Date	Pap test date	
SPEC_CASE	String	Special case	
UPD_OPER	Boolean	Update operator	
UPD_DATE	Date	Update date	
UPD_TIME	Time	Update time	HHMI(The format of all of the rest time in this document are the same as this one.)
TB_NO	String	Tuberculosis patient number	
FAX_AREA_NO	String	Fax area code	
FAX_NO	String	Fax number	Encrypted
BBC_NO	String	BB call/cellphone number	Encrypted
BEXPER_FLAG	String	Clinical trial	y:yes n: no
EVALUATE_DATE	Date	Evaluation date	
DEATH_DATE	Date	Date of death	D+ YYYMM
SPEC_DIES_FLAG	String	Special case	y:yes n: no
INSURRAN_FLAG	String	VIP patient	y:yes n: no
MEDIC_ACCID	String	Medical accid_flag	y:yes n: no
ACCID_INSUR	String	Acid_insur flag	y:yes n: no

Attributes	Type	Description	Values
REG_FLAG	String	Registration flag	y:yes n: no
DEAD_DATE	Date	Date of death	
VIP_FLAG	String	VIP patient flag	y:yes n: no
DEATH_PLACE	String	Place of death	
DEATH_ARCHIVE	String	Deceased patient archive	
DEATH_DESC	String	Cause of death	
DEATH_LOGIN_OPDER	String	Death registration personnel	

Table 6: Gemeli – Chemotherapy

Attributes	Type	Description	Values
Chemotherapy	Boolean	for patients who underwent chemotherapy	yes / no
Which chemo	String	name of chemotherapy schedule	GEM, FOLFOX, FOLFIRINOX, 5FU, NA
N° cycles	Integer	number of cycles of chemotherapy	
RT timing	String	radiotherapy timing	before chemo, after chemo, RT sandwich, only RT(CT)
Concomitant Chemo	Boolean	if the patient underwent concomitant chemotherapy	yes / no

5.2 Mapping to FHIR

After gathering the information from the pilots, the next step is to map the attributes to FHIR if possible, or to relevant ontologies.

FHIR is a description of how healthcare data can be organized and presented so that the meaning of the data is clear. Fundamental to the development of FHIR was the need to simplify how healthcare data can be exchanged between parties. As healthcare data becomes more widely digitized, and the interoperability between different healthcare systems becomes critical, FHIR's streamlined data model aims to help mitigate the compounding complexity of building robust healthcare data solutions.

In the context of iHelp, having a common data schema that most of the iHelp components can work on is crucial, as it allows the processing of data coming from different sources without the need of creating specialized processors for each one. The FHIR standard along with mCode, fit perfectly for this task as they provide many standardized medical terms that are used with the same meaning in most of the clinical organizations.

In order to extend the data provided by the pilots with the relevant mapping, in the data tables presented above, we add two new columns that holds the mapping and the link to the relevant source and another one holding the mapped values. Indicatively, we provide in the next table, a sample mapping done for attributes provided by Gemeli partner.

Attributes	Type	Description	Values	Links	FHIR mapping	FHIR Values
Sex	String		M, F	https://www.hl7.org/fhir/patient-definitions.html	Patient.gender	male, female
Age	Integer			https://www.hl7.org/fhir/person-definitions.html	*New field* if needed Person.birthDate can be used	
Ca19.9 diagnosis	Integer	marker at diagnosis		http://hl7.org/fhir/us/mcode/STU1/StructureDefinition-mcode-tumor-marker-definitions.html http://build.fhir.org/ig/HL7/fhir-mCODE-ig/	Observation.code Profile: TumorMarker TumorMarkerTest VS	possible options at: http://hl7.org/fhir/us/mcode/STU1/ValueSet-mcode-tumor-marker-test-vs.html
T	Integer	pathological staging T	1, 2, 3, 4	http://build.fhir.org/ig/HL7/fhir-mCODE-ig/StructureDefinition-mcode-tnm-primary-tumor-category-definitions.html	Observation.code Profile: TNMPrimaryTumorCategory Primary tumor.pathology Cancer	T1, T2, T3, T4
N	Integer	pathological staging N	0, 1, 2	http://build.fhir.org/ig/HL7/fhir-mCODE-ig/StructureDefinition-mcode-tnm-regional-nodes-category.html	Observation.code Profile: TNM Regional Nodes Category	N0, N1, N2

					Regional lymph nodes.pathology	
M	Integer	pathological staging M	0, 1	http://build.fhir.org/ig/HL7/fhir-mCODE-ig/StructureDefinition-mcode-tnm-distant-metastases-category.html	Observation.code Profile: TNM Distant Metastases Category Distant metastases.pathology [Class]	M0, M1
Tumor location	String		head, body, tail, isthmus, hooked process	http://build.fhir.org/ig/HL7/fhir-mCODE-ig/StructureDefinition-mcode-tumor.html	BodyStructure.location Profile: Tumor Cancer Body Location Value Set	
Surgery	Boolean	If the patient underwent surgery	yes / no	http://build.fhir.org/ig/HL7/fhir-mCODE-ig/StructureDefinition-mcode-cancer-related-surgical-procedure.html	*New field* if needed Procedure.code can be used with CancerRelatedSurgicalProcedure profile to define the exact procedure	yes/no

5.3 Next steps

As mentioned, the process of data gathering and analysis, takes place in iterations that involve discussions with the pilot partners in order to provide clarifications and the information needed, to comprehend, clean and map the data to FHIR. This is a time-consuming process, which is still on going the time this document is written, thus at this point, no “safe” information can be provided regarding the HHR model itself. Once the data gathering, analysis and validation is completed for each pilot, the next steps involve the definition of the relevant ontologies and the “merging” of this ontologies into the common HHR model. A more detailed description of the HHR model will be provided in the D3.2 “Data Modelling and Integrated Health Records: Design and open specification II”.

6 Conclusions

This deliverable gathers the available information for the data modelling and mapping approaches that will be utilized in iHelp. Based on the HHRs – one of the outcomes of the CrowdHEALTH project - the mapping procedure will be led from the Fast Healthcare Interoperability Resources (FHIR) specification, which is a standard for electronically exchanging healthcare information and data. The description and expansion of the HHR model, will be based on an ontological level in order to accomplish a high degree of semantic interoperability. The aim is to facilitate sharing of specific aspects of HHR, between different health platforms. For this reason, as a first step, the available data samples from the pilot cases where gathered and the next step is to map the attributes to FHIR if possible, or to relevant ontologies. A second version of this deliverable will be released on the later stages of the project, D3.2 "Data Modelling and Integrated Health Records: Design and open specification II", which will include the mapping between the data samples from the pilots available at that time to the HHR and it will contain additional information about possible extensions of the HHR model.

Bibliography

A. Begoyan, "An overview of interoperability standards for electronic health records", *USA: society for design and process science*, 2007.

X. Chen, D. Berry, and G. Stephens, "Using a generalised identity reference model with archetypes to support interoperability of demographics information in electronic health record systems", In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6834-6839, Aug. 2015.

ECOG-ACRIN. 2021. "ECOG Performance Status - ECOG-ACRIN". [Online] Available at: https://ecog-acrin.org/resources/ecog-performance-status [Accessed 01 August 2021].

N. Grønning, "Data management in a regulatory context. Frontiers in medicine", vol. 4, p. 114, 2017.

K. Häyrinen, K. Saranto, and P. Nykänen, "Definition, structure, content, use and impacts of electronic health records: a review of the research literature", *International journal of medical informatics*, vol. 77, no. 5, pp. 291-304, 2008.

Hl7.org. 2021. "Index - FHIR v4.0.1", Nov. 1, 2019. [Online] Available at: http://hl7.org/fhir/ [Accessed 03 August 2021].

Hl7.org. 2021. "FHIRPath (Normative Release)". [Online] Available at: http://hl7.org/fhirpath/ [Accessed 05 August 2021].

Hl7.org. 2021. "Change - FHIR v4.0.1". [Online] Available at: https://www.hl7.org/fhir/change.html [Accessed 03 August 2021].

K. Imachi, and T. Mussivand, "Outline of the International Organization for Standardization Standard for Circulatory Support Devices (ISO 14708-5)", *Artificial organs*, vol. 34, no. 9, pp. 695-698, 2010.

International Organization for Standardization, "ISO 12967:2009, Health informatics—Service architecture", 2009.

International Organization for Standardization, "ISO 14155, International Standard for Clinical Investigation of Medical Devices". Geneva, Switzerland, 1996.

International Organization for Standardization, "ISO/TS 22220:2011, Health informatics -- Identification of subjects of health care", 2011.

Y. A. Jung, and Y. J. Kim, "Comparative study of ISO standards for an effective implementation of the domestic medical device GMP system", *Journal of the Korean Society for Quality Management*, vol. 46, no. 2, pp. 211-224, 2018.

G. O. Klein, P. A. Sottile, and F. Endsleff, "Another HISA-the new standard: health informatics-service architecture", In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*, p. 478, 2007.

K. Koshechkin, G. Lebedev, and F. Eduard, "Implementation of IDMP Standards as a Means of Creating a Unified Information Space in the Field of Drug Circulation", *Procedia Computer Science*, vol. 176, pp. 1745-1753, 2020.

P. Kubben, M. Dumontier, and A. Dekker, "Fundamentals of clinical data science", 2019.

D. Kyriazis et al, “CrowdHEALTH: Holistic Health Records & Big Data Analytics for Health Policy Making and Personalized Health”, *Informatics Empowers Healthcare Transformation*, 238, p. 19, 2017.

C. Martínez-Costa, M. Menarguez-Tortosa, and J. T. Fernández-Breis, “Towards ISO 13606 and openEHR archetype-based semantic interoperability”, In *Medical Informatics in a United and Healthy Europe*, pp. 260-264, 2009.

Mcodeinitiative.org. 2021. “mCODE | Minimal Common Oncology Data Elements”. [Online] Available at: https://mcodeinitiative.org/ [Accessed 03 August 2021].

Medlineplus.gov. 2021. “MedlinePlus: Understanding Medical Words Tutorial”. [Online] Available at: https://medlineplus.gov/medwords/medicalwords.html [Accessed 05 August 2021].

P. Muñoz, J. D. Trigo, I. Martínez, A. Muñoz, J. Escayola, and J. García, “The ISO/EN 13606 standard for the interoperable exchange of electronic health records”, *Journal of Healthcare Engineering*, vol. 2, no. 1, pp. 1-24, 2011.

Omg.org. 2021. “About the Unified Modeling Language Specification Version 2.0”. [Online] Available at: https://www.omg.org/spec/UML/2.0/About-UML/ [Accessed 30 August 2021].

V. A. Perkins, “IDMP: AN INTERNATIONAL STANDARD FOR LIFECYCLE DATA MANAGEMENT”, *Applied Clinical Trials*, vol. 26 (4/5), p. 10, 2017.

M. D. Santos, M. P. Bax, and D. Kalra, “Dealing with the archetypes development process for a regional EHR system”, *Applied clinical informatics*, vol. 3, no. 03, pp. 258-275, 2012.

W. V. Slack, and C. W. Slack, “Patient-computer dialogue”, *New England Journal of Medicine*, vol. 286, no. 24, pp. 1304-1309, 1972.

S. Schulz, R. Stegwee, and C. Chronaki, “Standards in healthcare data”, *Fundamentals of Clinical Data Science*, pp. 19-36, 2019.

P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, and D. Z. Sands, “Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption”, *Journal of the American Medical Informatics Association: JAMIA*, vol. 13, no. 2, pp. 121–126, 2006, doi: 10.1197/jamia.M2025.

TC251, C. E. N., “Towards concurrent use of ContSys, 13606, and HISA”, In *Report from 2nd Workshop, Madrid,* vol. 7, p. 2013, 2013.

J. Werner, S. E. Combs, C. Springfeld, W. Hartwig, T. Hackert, and M. W. Büchler, “Advanced-stage pancreatic cancer: therapy options”, *Nature reviews Clinical oncology*, vol. 10, no. 6, pp. 323-333, 2013. doi: 10.1038/nrclinonc.2013.66.

Annex A: The CrowdHEALTH HHR conceptual model

In this section we are presenting the HHR conceptual model as it was defined in the CrowdHEALTH project. The intention in the iHelp project is to use this work as the baseline and evaluate if it possible to reuse at least part of it, so the purpose of the section is to provide the reader with an overview of this conceptual model as a potential "sneak peek" of the future work.

The conceptual HHR data model represents structured health records including several types of information that were relevant to the patient's health. HHRs were described using UML class diagrams, which are divided in several packages aiming at facilitating the representation and the description of the reported information. Each package contains details related to a specific subject, such as clinical Conditions of patients or Measurements, reporting all entities and their reciprocal relationships.

This section aims at summarising the fragments of the HHR model defined in the CrowdHEALTH project. Starting from the reported work, the HHR model can be extended and updated to satisfy the requirements collected from the involved pilots in the iHelp project.

The UML packages of the HHR model, detailly described in the following sections, were:

- Identifier
- Person
- Organization
- Condition
- Activity
- Episode of care
- Measurement
- Quantitative measures
- Laboratory tests measures

Primitive data types defined by the HHR model are shown in Figure x. Their semantics is the same of the FHIR data types.

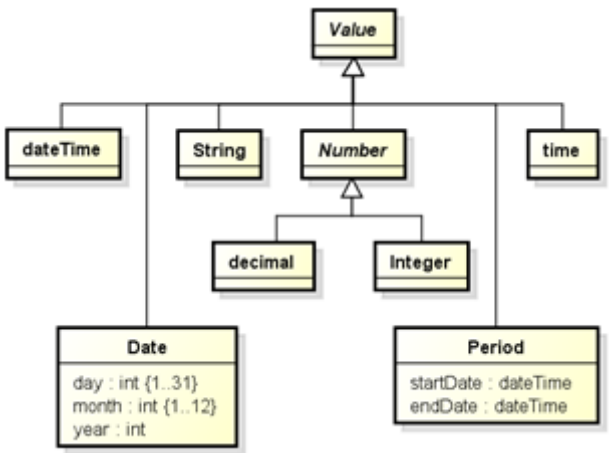

Figure x: HHR Primitive data type model

6.1.1.1 Identifier

All entities of the HHR model inherit from IdentifiedEntity, i.e. any entity can be identified with a string id that is unique within a given IdentifierSystem. An IdentifiedEntity has at least one Identifier denoting a numeric or alphanumeric string associated with a single entity within a given identifier system, and each identifier is generated by one system. The systems admitted in the HHR model are listed in the coded class IdentifierSystem, that represents a standard to connect a unique id to each entity belonging to a specific context. Each identified entity may have only one identifier per IdentifierSystem: two identified entities cannot share the same identifier belonging to the same IdentifierSystem.

6.1.1.2 Person

This package qualifies a person or a group of people. The class Person represents demographics and administrative information of a person regardless of a specific health context. The Gender enumeration models the gender while the class Address, having a specific AddressUse and AddressType, models a person's address and birthplace.

Class Group denotes a group of people with a common set of characteristics. The reification of a valorised property/attribute is represented by the class Characteristic while the CharacteristicType is the attribute/property that is reified by a characteristic. If the identity of the group members is not known, the Group is specialized in AnonymisedGroup. Person and Group inherit from PersonOrGroup that, as well as AutomaticAgent, inherit the unique identifier from their superclass Agent.

Class Coverage, associated to Person, models an insurance plan, whose status is determined with the FinancialResouceStatus enumeration.

The same person can play different individual roles within different contexts. Each individual role of the same person is represented by a different instance of the class PersonInTime, that is a view of a person related to a specific time frame and/or a specific context (denoted by the specific subclass). The same person can be linked to several instances of PersonInTime, each of which contains information specific to the corresponding role and is connected to the person playing the role by the "player" association-end.

Specifically, a person has the role Patient when subject to the health care activities provided by HealthCare professionals; a person assisted by two different healthcare providers plays two different Patient roles. A qualified medical doctor employed in a specific organization person has the role of Practitioner, with one or more PractitionerSpecialty; a practitioner working for different organizations plays different instances of PractitionerRoleType, matching the set of the roles carried out at an organization for a period. HealthCarePerson, the superclass of Patient and Practitioner, represents any person playing a role in an HealthCareOrganization (see section 4.4.1.3).

When a Patient dies, the cause of the death is represented by an instance of the class ConditionType.

6.1.1.3 Organization

A MedicalDepartement belongs to one and only one CenterDepartment, made of different medical departments. MedicalDepartments are (subclasses of) a NestedOrganization, because they can be part of another organization. The superclass of NestedOrganization, Region and CenterDepartement is Organization, that models a formally or informally recognized grouping of

people or organizations constituted with the goal of realizing some form of collective action, including companies, institutions, corporations, departments, and so on.

Details and location of a physical place where services are provided and resources and participants may be stored, found, contained, or accommodated can be found in the class `Location`.

6.1.1.4 Condition

A `PersonOrGroupEvent` is the record of an automatic or manual activity or observation or request of an activity, performed on a person or group (e.g. the administration of a medication to a patient) or generating information about a person or group (e.g. the calculation of the BMI of a person); the event should not necessarily be related to a healthcare service. If the event is a request to perform an activity, then its attributes refer to the requesting action and not to the requested activity; for instance, the performer of a `MedicationRequest` is different from the performer of the `MedicationAdministration` (see section 4.4.1.5).

The HHR model introduces various types of events:

- A `Condition` is a statement about an objective state of a patient, done by the patient itself, a group of persons or a practitioner. Condition refers to a persistent state and is distinct from a `PersonOrGroupObservation` (see Measurement section) that instead refers to a particular instant in time. Condition has two subclasses: `ClinicalFinding` and Diagnosis, having an association-end "type" with `ClinicalFindingType` and `DiagnosisType` respectively, both subclasses of `ConditionType`. An instance of `ClinicalFindingType` represents a statement about a persistent objective status of a patient while a `DiagnosisType` is a statement that is the outcome of a cognitive process, i.e. it is the understanding of a set of measures and/or clinical findings. The status of the clinical condition of the `PersonOrGroup` may be one of the values in the `ClinicalStatus` enumeration. `BodySite` represents the anatomical location where the Condition occurs while `Stage` specifies its stage or grade and is specialized in `ChronicStage` when the condition becomes chronic.
- `PersonOrGroupObservation` represents the evidence at the base of the `Condition`. The severity association specifies the current severity of the patient's condition and may have one of and may be one of the values in the `Severity` enumeration.

6.1.1.5 Activity

All activities included in the HHR model, either planned or performed, are instances of `PersonOrGroupEvent`. Each event has an `EventStatus`, and each specialization of `PersonOrGroupEvent` may use a specialized set of status. More instances of `PersonOrGroupEven` can be grouped in a `CollectionOfEvents` having an instance of Group as subject. The category of the members belonging to the collection of events can be set with the `PersonOrGroupCategory` enumeration.

A `Procedure` is a specialization of `PersonOfGroupEvent` and represents any action done on a person while its specialization, `ClinicalProcedure`, represents any medical procedure performed within a hospital/clinic. The type of the medical procedure is specified with the coded class `ClinicalProcedureType`. Procedure is associated with `Encounter` (see section 4.4.1.6) representing the encounter during which the procedure was performed. `ClinicalProcedure` and

ClinicalProcedureType are linked to ConsultationORTreatmentEncounter (see section 4.4.1.6). The ProcedureStatus enumeration allows to describe the status of a procedure and can assume a predefined and finite number of different values.

MedicationEvent is an event representing the administration of a medication. Each MedicationEvent is associated to a Medication, i.e. the substance that is used as an administered medication, produced by the hospital pharmacy or by a brand. The composition of a Medication is specified by one or more instances of the class Ingredient where the player is a MedicationOrSubstance. The recommended dosage is represented by the association recommendedDosage with the class Dosage, that is related to Timing and ContinuousDimensionfulQuantity: Timing specifies an event that may occur multiple times inside a DurationUnit time frame while ContinuousDimensionfulQuantity represents the amount of medication per dose. A Substance is a pure substance or a homogeneous mixture with a definite composition, specified by zero or more ingredients, i.e. other substances. MedicationEvent is specialized by MedicationStatement and MedicationAdministration. MedicationStatement represents an assertion, by some individual (the asserter), that a specific medication has been prescribed (but not necessary taken) or has been taken (also without a prescription). MedicationAdministration represents an administration of a medication asserted by an individual employed in the organization providing it. MedicationStatementStatus and MedicationAdministrationStatus respectively represent their status. A MedicationRequest is an order or request for both supply of the medication and the instructions for administration of the medication to a patient/person, authorized by a performer with a specific PractitionerSpecialty. The enumeration MedicationPrescriptionIntent models the intent of the prescription while the association performerOrganization describes the Organization of the Practitioner that performed the MedicationRequest.

a MedicationDispense indicates that a medication product is to be or has been dispensed for a named person/patient. This includes a description of the Medication and the instructions for administering it. The MedicationDispense is the result of a pharmacy system responding to a MedicationRequest order. Its status is represented by the DispenseStatus enumeration.

A CarePlan describes how one or more practitioners intend to deliver care for a particular patient, group or community for a period of time, possibly limited to care for a specific condition or set of conditions. CarePlan represents both proposed plans and active plans. Its status may be one of the values in the CarePlanStatus enumeration. The association “intend” with CarePlanIntent indicates the level of authority/intentionality associated with the care plan and where the care plan fits into the workflow chain while the CarePlanType specifies the kind of the plan (e.g. "home health", "psychiatric", "asthma", "disease management", "wellness plan", etc.).

6.1.1.6 EpisodeOfCare

The class EpisodeOfCare represents a set of health care activities to manage a specific issue. A EpisodeOfCare can consist of zero or more instance of Encounter, i.e. events representing an interaction between a patient (or a group of people) and healthcare provider(s) with the purpose of providing healthcare service(s) or evaluating the health status of a patient. The “service” association between Encounter and the coded class EncounterServiceType (e.g. e-mail consultation, surgical day-care, skilled nursing, rehabilitation) allows to specify the type of the encounter. Encounter registers the details of an activity directly related to the patient, while EpisodeOfCare is the container that can connect

a series of Encounter for problems/issues. During an encounter different events can occur, therefore Encounter has an association 0..* with the abstract class PersonOrGroupEvent representing a generic event. MedicalDepartment is the organization in charge of maintaining the information of an Encounter. The urgency of the encounter is represented with the enumeration Priority.

The class Appointment is used to reserve a healthcare event among patients, practitioners and/or related persons for a specific date/time and can result in one or more Encounter(s). ClinicalSpecialty represents the specialty of the practitioner needed to perform the service required in the Appointment while the RequestType is the type of request for which the appointment is set.

The status of EpisodeOfCare, Encounter and Appointment is represented by the enumerations EpisodeOfCareStatus, EncounterStatus and AppointmentStatus.

Location is the physical place where services are provided while the coded class LocationType represents the type of the Location from the functional point of view. Encounter has four associations with Location:

- "admissionLocation": the location where the admission of the patient is performed;
- "dischargeLocation": location from where the patient is discharged;
- "destination": location to which the patient will move from the discharge location;
- "location": list of locations where the patient has been during this encounter.

Each Encounter is justified by a reason. Possible reasons are listed in the EncounterReason enumeration; HospitalizationReason represent reasons to be admitted to inpatient encounter and HospitalizationReason denote reasons to be admitted to an emergency.

The DischargeDisposition enumeration represents the location/status following the encounter (e.g. death, transfer to home/hospice/snf/AMA); possible discharge dispositions after an hospitalization encounter are represented by HospitalizationDischargeDisposition and possible discharge dispositions for an emergency encounter are represented by the EmergencyDischargeDisposition.

Encounter has several specializations:

- HospitalizationEncounter represents an inpatient encounter where a patient is admitted to a hospital to receive a specific healthcare service. AdmitSource represents from where the patient came (e.g. physician referral, transfer etc.) while DischargeReason represents the reason for the discharge.
- EmergencyEncounter represents an encounter where the patient receives immediate evaluation and treatment, provided until he or she is discharged or the responsibility for the patient's care is transferred elsewhere.
- HospitalAtHomeEncounter happens in residence of the patient belonging to one of the categories listed in the enumeration PatientType. PrimaryCareEncounter represents a primary care consultation of the target patient that takes place within the office of a primary care physician.
- ConsultationORTreatmentEncounter is an abstract class representing a formal meeting, typically not requiring a creation of a medical record, with a medical doctor/healthcare operator to observe/monitor/discuss/ask advice/undergo ambulatory treatments. The type of the

consultation or treatment encounter can be set choosing one of the values listed in the enclosed enumeration CTEType. The association between ConsultationORTreatmentEncounter and ClinicalProcedureType represents the types of any clinical procedure performed during this encounter. ConsultationORTreatmentEncounter can be specialized into 2 subclasses:

- InpatientEncounter: an encounter occurring within a hospital facility, requiring the patient to stay for the night.
- OutpatientEncounter: an encounter taking place within a hospital facility, not requiring the patient to stay for the night.

6.1.1.7 Measurement

A PersonOrGroupObservation is a simultaneous observation or assertion of the value of one or more (e.g. as with ComposedMeasure) properties about a Person or about an entity related to a Person, done by the person itself, a Device or some other actor.

PersonOrGroupMeasure is a specialization of PersonOrGroupObservation, modelling a measure performed according to a standardized method whose results can be compared to expected min and max values for a normal (i.e. healthy) person. It is related to an ObservationStatus, representing the status of the measurement observation, to an Interpretion enumeration, representing an evaluation based on the result of the observation, to a MeasureMethod, representing the procedure followed to perform the measurement, and to a Device, representing the device used to generate the observation data. The measure can be a ComposedPersonOrGroupMeasure or an AtomicPersonGroupMeasure: the former represents the simultaneous measurement of the value of two or more properties about a Person or AnonymisedGroup (e.g. the measurement of the blood pressure, embodying systolic and diastolic pressure) and the latter a measurement of the value of a single property about a Person or AnonymisedGroup. An AtomicPersonGroupMeasure is linked to the class NormalRange representing a range of the values of a typical normal (i.e. healthy) PersonOrGroupMeasure, generally depending on clinical guidelines and measurement method.

AtomicPersonGroupMeasure is specialized by:

- BooleanPersonOrGroupMeasure, modelling the measurements where the value can be true or false (e.g. to state if a person is or is not a smoker).
- CategoricalPersonOrGroupMeasure, that represents the measurement of value belonging to a certain Category.
- QuantitativePersonOrGroupMeasure represents the measurement of a Quantity which magnitude is represented by a number. It is further specialized in VitalSignMeasure and PhysicalExamOrTest. VitalSignMeasure represents a clinical observation that measures a body's basic functions such as blood's basic functions or the blood pressure, heart rate, respiratory rate, pulse oximetry etc. PhysicalExamOrTest represents a physical examination (e.g. measure or BMI) or a physical performance test (e.g. time needed to run 60 meters).
- The class Value is specialized in Range, Quantity, BooleanValue and Category. Quantity values are further specialized in ContinuousQuantity, in which the magnitude is represented by a real number and DiscreteQuantity in which the magnitude is represented by an integer. ContinuousQuantity values are specialized in ContinuousDimensionfulQuantity in which the magnitude has a specific unit of measure and ContinuousDimensionlessQuantity

in which the value does not have any unit of measure. `DiscreteQuantity` has three specializations: `CardinalityQuantity`, `PercentileQuantity` and `QuartersQuantity`. The magnitude of `CardinalityQuantity` is a dimensionless discrete integer quantity while the magnitude of `PercentileQuantity` is an integer included between 0 and 100.

Class `ContinuousDimensionfulQuantity` has many specializations bound to a unit of measure. For instance, a `TempoQuantity` is a continuous quantity which value is a real number where the unit of measure can assume one of the values listed in the `TempoUnit` enumeration. This logic is applied to all subclasses of `ContinuousDimensionfulQuantity`.

Class `ConcentrationQuantity` is specialized by `SubstanceConcentrationQuantity` and `RatioQuantity`: the first one representing the result of a measure intended as ratio between two different units (e.g. g/dL, mg/dL, nmol/L, umol/L etc.) while the second one representing any value calculated as the ratio between a quantity X and a quantity Y having the same units of measurements (expressed as parties per ten, or percentage etc.).

6.1.1.8 Quantitative measures

Quantitative measures are measurements of values of type `Quantity`, which have a (numerical) magnitude and optionally a unit of measure. This kind of measurements is specialized by:

- laboratory test measures
- fitness measures
- heart rate and blood pressure measures
- lung tests measures
- lifestyle measures
- sleep measures
- social measure
- demographic measures
- workout measures

6.1.1.8.1 Laboratory tests measures

`LaboratoryTest` measures are quantitative measures representing the results of laboratory analytic services in areas such as chemistry, haematology, serology, histology, cytology, anatomic pathology, microbiology, and/or virology. Laboratory tests included in the HHR model are:

- `C_ReactiveProtein`: the c-reactive protein measurement.
- `GlycosilatedHemoglobin`: the glycohemoglobin measurement.
- `UrineAlbumin`: the urine albumin measurement.
- `Glucose`: the glucose measurement.
- `BloodUrea`: the blood urea measurement.
- `Creatinine`: the creatinine measurement.
- `TotalCholesterol`: the total cholesterol measurement.
- `LowDensityCholesterol`: the low-density lipoprotein cholesterol measurement.
- `HighDensityCholesterol`: the high-density lipoprotein cholesterol measurement.
- `GPTTransiminases`: the alanine aminotransferase measurement.

- `SerumAlbumin`*:* the serum albumin measurement.
- `Calcium`: the calcium measurement.
- `Sodium`: the sodium measurement.
- `Potasium`: the potassium measurement.
- `TransferrineSaturationIndex`: the transferrin saturation index.
- `Ferrinite`: the ferritin measurement.
- `Transferrine`: the transferrin measurement.
- `ArterialPh`: the haemoglobin free measurement.
- `Hematocrite`: the platelet haematocrit measurement.
- `FreeT4`: the T4 free measurement.
- `FreeT3`: free tri-iodothyronine measurement used to assess thyroid function.
- `VenousPh`: the ph measurement venous.
- `PROBNP`: the prohormone of brain natriuretic peptide measurement.
- `ArterialO2`: the arterial O2 measurement.
- `Troponin_T`: the troponin T measurement.
- `Haemoglobin`: the haemoglobin measurement.
- `ArterialCO2`: the arterial CO2 measurement.
- `TotalT4`: the total T4 test.
- `TotalT3`: the toalT3 test.
- `VenousO2`: the venous oxygen saturation measurement.
- `Triglycerides`: the triglycerides measurement.
- `AlbuminCreatinineRatio`: the albumin creatinine ratio measurement.
- `VenousCO2`: venous-to-arterial carbon dioxide measurement.
- `PeripheralOxygenSaturation`: the peripheral oxygen saturation measurement.
- `HighSensitivityCReactiveProtein`*:* high sensitivity c reactive protein test.
- `Phosphate`*:* the phosphate measurement.
- `ProteinKinase`*:* the protein kinase measurement.
- `GOTTransaminase`*:* the GOT transaminases measurement.
- `TSHThyroid`*:* the TSH thyroid measurement.
- `Prolactine`*:* the prolactin measurement.
- `Follitropin`*:* the follitropin measurement.
- `Lutropin`*:* the lutropin measurement.
- `Progesterone`*:* the progesterone measurement.
- `Estradiol`*:* the oestradiol measurement.
- `Testosteron`*:* the testosterone measurement.
- `DehydroepiandrosteroneSulfate`*:* the dehydroepiandrosterone sulphate measurement.
- `Dehydroepiandrosterone`: the dehydroepiandrosterone measurement.
- `17-Hydroxyprogesterone`: the 17 hydroxyprogesterone measurement.
- `Androstenedione`: the androstenedione measurement.
- `SexHormoneBindingGlobulin`*:* the sex hormone binding globulin measurement.
- `Corticotropin`*:* the corticotropin measurement.
- `Choriogonadotropin`*:* the choriogonadotropin measurement.

- `PlasmaReninConcetration`*:* the plasma renin concentration measurement.
- `Cortisol`*:* the cortisol measurement.
- `Aldosterone`*:* the aldosterone measurement.
- `C-Peptide`*:* the C-peptide measurement.
- `Somatotropin`*:* the somatotropin measurement.
- `Insulin`*:* the insulin measurement.
- `Parathyroid`*:* the parathyroid measurement.
- `Osteocalcin`*:* the osteocalcin measurement.
- `Calcitonin`*:* the calcitonin measurement.
- `Calcitriol`: the calcitriol measurement.
- `Calcidiol`: the calcidiol measurement.
- `ChoriogonadotropinBeta`: the choriogonadotropin beta measurement.
- `Homocysteine`: the homocysteine measurement.
- `Alpha1Fetoprotein`: the alpha 1 fetoprotein measurement.
- `Thyroglobulin`: the thyroglobulin measurement.
- `ErythrocyteSedimentation`*:* the erythrocyte sedimentation measurement.
- `APTTInBloodByCoagulationAssay`*:* APTT in blood by coagulation assay measurement.
- `ProthrombinTime`*:* the prothrombin time measurement.
- `Cobalamin`*:* the cobalamin measurement.
- `Urate`*:* the urate measurement.
- `FibrinogenInPlateletPoorPlasma`*:* the fibrinogen in platelet poor plasma measurement.
- `Lipoprotein`*:* the lipoprotein measurement.
- `Magnesium`: the magnesium measurement.
- `Folate`: the folate measurement.
- `Iron`: the iron measurement.
- `BilirubinTotal`: the bilirubin total measurement.
- `BilirubinDirect`: the bilirubin direct measurement.
- `Leukocytes`: the leukocytes measurement.
- `AspartateAminotransferase`: the aspartate aminotransferase measurement.
- `AlanineAminotransferase`: the alanine aminotransferase measurement.
- `GammaGlutamylTransferase`: the gamma glutamyl transferase measurement.
- `AlkalinePhosphatase`*:* the alkaline phosphatase measurement.
- `CreatineKinase`*:* the creatine kinase measurement.
- `Amylase`: the amylase measurement.
- `ABOGroupBlood`: the group blood measurement, it can assume one of the values listed in the enclosed *ABOGroupBloodCategory* enumeration.
- `RhBlood`: the RH blood measurement, it can assume one of the values listed in the enclosed enumeration *RhBloodCategory*.
- `Hormone`: the measure of the hormone;
- `Cystatin-c`*:* the measure of cystatin-c;
- `DipstickAlbuminuria`*:* the measure of dipstick albuminuria;
- `Bicarbonate`*:* the measure of the bicarbonate;

- `BloodLipids`*:* the measure of the blood lipids;
- `CarbonDioxide`*:* the measure of the carbon dioxide.

6.1.1.8.2 Fitness measures

Fitness measures are quantitative measures representing the set of parameters measured during a fitness test, and physical body parameters. Specifically, fitness measures are:

- `Waist`: the length of the waistline of the person.
- *Height*: the height of the person.
- `Weight`: the weight of the person.
- `TricepsSF` (Triceps skin fold): the width of a fold of skin taken over the triceps muscle. Triceps skinfold reflects the amount of peripheral sub cutaneous fat. It is a proxy for body composition.
- `Run600m`: the time needed to run 600 meters.
- `Dash60m`: the time needed to run 60 meters.
- `StandAndReach` (flexibility of lower back and legs): the measure of the distance fingertips reaches past the toes during a stand-and-reach test.
- `BMI`: body mass index of the person, defined according to World Obesity Federation (represented by a value of the *WOF_CODE* enumeration).
- `PolygonBackwards`: the time needed to complete a 10m distance during the polygon backwards fitness test. It is measured during a fitness test designed to measure coordination. The subject moves backwards on all fours and covers a 10m distance. The subject craws over and under the 35 cm high obstacles that are placed at 3 meters and 6 meters from the starting line, respectively.
- `ArmPlateTapping`: the number of taps completed in 20 seconds. It is measured during a reaction test using a tapping action which measures upper body reaction time, hand-eye quickness and coordination. It is designed to assess the speed and the coordination of limb movement.
- `StandingBroadJump`: the distance jumped from a standing position. This test measures explosive leg power.
- `SitUp60s`: the number of sit-ups performed in 60 seconds. It is a measure of the strength of the trunk.
- `BentArmHang`: the time that a person can hold a flexed arm hang position above a horizontal bar.
- `HR_FI`: the summary measure of health-related fitness. It is calculated as the sum of individual z-scores from 3 motor tests related to health (tests that assess endurance and muscular strength and muscular endurance, i.e. 600m run, sit-ups and bent arm hang).
- `PR_FI`: the summary measure of performance-related fitness. It is calculated as the sum of individual z-scores from 4 motor tests related to performance (tests that assess explosive strength, agility and speed i.e. standing broad jump, 60m run and polygon backwards and arm-plate tapping).
- `Total_FI` (total physical fitness index). The sum of the individual z-scores of 8 different motor tests: `ArmPlateTapping`, `PolygonBackwards`, `BentArmHang`, `SitUp60s`, `StandingBroadJump`, `Dash60m`, `Run600m` and `StandAndReach`.
- `NumberOfWalkedSteps`*:* the number of steps walked.
- `BodyFatPercentage`*:* the percentage of body fat.

6.1.1.8.3 Heart rate and blood pressure measure

`HeartRateMeasure` is a quantitative measure representing the heart rate of an individual. Its value is an `HeartRateQuantity`, which has the unit of measure of `HeartBeatUnit`, and it is a specialization of `TempoQuantity`. `BloodPressure` is a measure composed of `SystolicBloodPressure` and `DiastolicBloodPressure`, which are two quantitative measures inheriting from `PressureMeasure`. `PerfusionIndexTissueByPulseOximetry` represents the perfusion index tissue by pulse oximetry of an individual, which value is continuous dimensionless quantity. `PulseOximetryWaveform` represents the measurement of the pulse oximeter.

6.1.1.8.4 Lung tests measures

Lung tests are some exams aiming at assessing the health status of the lungs of a person. In particular, the tests included in the HHR model, specializations of `PhysicalExamOrTest`, are:

- `ForcedExpiratoryFlowBetween25and75VitalCapacit`*:* Forced expiratory flow rate between 25-75% of vital capacity;
- `ForcedExpiredVolumeIn1Second`*:* Forced expired volume in 1 second;
- `ForcedVitalCapacity`*:* Forced vital capacity;
- `ForcedExpiratoryVolume1SecondVitalCapacityPercent`*:* Forced expiratory volume in one second/Forced vital capacity percent*;*
- `ForcedExpiredVolumeIn6Seconds`*:* Forced expired volume in 6 seconds;
- `PeakExpiratoryFlowRate`*:* Forced expired volume in 6 seconds.

6.1.1.8.5 Lifestyle measures

<u>Alcohol and tobacco consumer model</u>

`IsAlcoholConsumer` and `IsTobaccoConsumer` classes, specializations of `BooleanPersonOrGroupMeasure`, denote whether a person or a group of persons are respectively alcohol consumers or smokers.

<u>Active physical activity model</u>

Classes representing the active lifestyle physical activities data of a person or group of people indicate the percentage of time (e.g. three minutes per day) when the subject has been physically active. The related information is:

- `ModerateToVigorousPhysicalActivityWeekdaysAverage`: moderate-to-vigorous physical activity on school days (Monday to Friday), average time per period (e.g. average minutes per day);
- `ModeratePhysicalActivityWeekdaysAverage`: moderate physical activity on school days (Monday to Friday), average time per period (e.g. average minutes per day);
- `ModerateToVigorousPhysicalActivityWeekendAverage`: moderate-to-vigorous physical activity on weekend days (Saturday–Sunday), average time per period (e.g. average minutes per day);
- `VigorousPhysicalActivityWeekdaysAverage`: vigorous physical activity on school days (Monday to Friday), average time per period (e.g. average minutes per day);

- `ModeratePhysicalActivityWholeWeekAverag`: moderate physical activity on whole week (Monday–Sunday), average time per period (e.g. average minutes per day);
- `VigorousPhysicalActivityWholeWeekAverage`: vigorous physical activity on whole week (Monday–Sunday), average time per period (e.g. average minutes per day);
- `VigorousPhysicalActivityWeekendAverage`: vigorous physical activity on weekend days (Saturday–Sunday), average time per period (e.g. average minutes per day);
- `ModeratePhysicalActivityWeekendAverage`: moderate physical activity on weekend days (Saturday–Sunday), average time per period (e.g. average minutes per day);
- `ModerateToVigorousPhysicalActivityWholeWeekAverage`: moderate-to-vigorous physical activity on whole week (Monday–Sunday), average time per period (e.g. average minutes per day) which interpretation can assume one of the values listed in the enclosed `PhysicalActivityLevel` enumeration.

Sedentary physical activity model

Classes denoting the sedentary lifestyle physical activities data of a person or group of people show the percentage of time (e.g. three minutes per day) when the physical activity of the subject has been sedentary. The information provided are:

- `ScreenTimeWeekdays`: screen time on school days, average time per period (e.g. average minutes per day);
- `ScreenTimeWeekend`: screen time on weekend days, average time per period (e.g. average minutes per day);
- `ScreenTimeWholeWeekAdequacy`: screen time on whole weekdays, average time per period (e.g. average minutes per day) which interpretation can assume one of the values listed in the enclosed `ScreenTimeWholeWeekAdequacy` enumeration.

Summary physical activity model

Summary reports summarise physical activities in terms of energy fat burn and distance travelled collected by the wearables. They are composed of a set of information. Classes representing information related to fat burn, each one specializing DayVitalSignMeasure, are:

- `DayEnergyBase`: aka idle energy;
- `DayEnergyFatBurn`: the total energy burned in a day while in fat burn zone (where the subject had 50-69% of maximal heart rate);
- `DayAdditionalEnergyExpenditure`: the total consumed energy in a day in addition to the `DayEnergyBase` value.
- `DayEnergyCardio`: the total energy burned in a day while doing cardio (where the subject had 70-84% of maximal heart rate).
- `DayEnergyPeak`: the total energy burned in a day while in peak performance (where the subject had more than 85% of maximal heart rate).
- `DayEnergyOutOfFatBurn`: the total energy burned in a day while out of the fat burn zone (where the subject had less than 50% of maximal heart rate).

Classes representing the distance travelled in a day, specializing DayActivityMeasure, are:

- `DayNumberOfFloorsClimbed`: the total number of floors climbed in a day;
- `DayDistanceVeryActive`: the total distance travelled in a day while the subject was very active (according to the definition of activity levels used by Fitbit);
- `DayDistanceFairlyActive`: the total distance travelled in a day while the subject was fairly active (according to the definition of activity levels used by Fitbit);
- `DayDistanceSedentaryActive`: the total distance travelled in a day while the subject was sedentary active (according to the definition of activity levels used by Fitbit);
- `DayDistanceLightlyActive`: the total distance travelled in a day while the subject was lightly active (according to the definition of activity levels used by Fitbit);
- `DayAverageElevation`: the average elevation (i.e. altitude) where the activities of a day were performed.

Social data model

`SocialMeasure` takes into account how a person or a group of people interact in social contexts by including the following social data:

- `AverageCallDurationInFrame`*:* average duration of calls performed in the time period specified by the performedWhen attribute.
- `CallsTotalDurationInFrame`*:* total duration of all calls performed in the time period specified by the performedWhen attribute.
- `AverageEmergencyCallDurationInFrame`*:* average duration of emergency calls performed in the time period specified by the performedWhen attribute.
- `EmergencyCallsTotalDurationInFrame`*:* total duration of emergency calls performed in the time period specified by the performedWhen attribute.
- `EmergencyCallsInFrame`: the number of emergency calls performed in the given time frame.
- `CallsInFrame`: number of calls performed in the time period specified by the performedWhen attribute.
- `SocialContactsCount`*:* number of social contacts (friends, relatives, etc.) of the user at the time period specified by the performedWhen attribute.
- `MultimediaIItemsCount`*:* number of multimedia items (pictures, videos, etc.) of the user.

Sleep measurement model

`SleepSessionDuration` measures the duration of an entire sleep session. A single sleep session consists of a set instances of `SleepTypeMeasure` representing the type of sleep in a specific frame of the sleep session which belongs to. `SleepTypeMeasure` can assume one of the values listed in the enclosed `SleepTypeCategory` enumeration.

Classes related to the average sleep time per day during the days schools (Monday to Friday) and weekend are:

- `SleepWeekdaysAverage`: sleep (in the bed) time on school days (Monday to Friday), average time per period (e.g. average minutes per day);
- `SleepWeekendAverage`*:* percentage of sleep (in the bed) time on weekend days, average time per period (e.g. average minutes per day)

- `SleepAdequacy`: quality of sleep within the measured period. It can assume one of the values listed in the enclosed `SleepTimeAdequacy` enumeration.

Food intake

Each instance of the abstract class `FoodIntake` represents the observation of a specific event of food consumptions performed by a specific Person and inherits all `PersonOrGroupObservation` properties to identify the subject of the information, and who performed the observation (the performer), who stated (the asserter) and stored (the recorder) the information and when. Three kinds of `FoodIntake` are considered: an entire Meal, a single Course and a specific `FoodPortion`. A Meal may be of different types (e.g. breakfast or lunch) and may be of different Courses, where each Course represents the consumption of a set of `FoodPortions`. The property `declaredFoodQuantity` of the class `FoodPortion` represents the quantity of consumed food as stated by the asserter of the information (that is typically the consumer of the food). The food quantity can be provided by stating a weight or a different type of measure, e.g. a glass of orange juice; in this case some conversion table is used.

A food is in general composed of different nutrients. `SubstanceIntakeAmount` represents the amount of content of a specific substance contained in the specific portion of consumed food. Depending on the applications, it is possible to use such the property `estimatedSubstanceIntakeAmount` to express an estimate of the composition of the consumed food (e.g. cheese cake), with respect to a substance that is a nutrient (e.g. protein) or a micro-nutrient (e.g. tocopherol) or a food category (e.g. dairy) or another food (e.g. quark cheese).

Food intake frequencies

`FoodOrBeverageIntakeFrequency` is an abstract class which instances represent categorical measures about the frequency of consumption of a specific food and is specialized by:

- `FoodIntakeFrequency` represents the frequency of consumption of specific foods
- `DrinkingFrequency` represents the consumption of water or other hydrating beverage.

Allergy intolerance model

`AllergyIntolerance`, subclass of PersonOrGroupEvent, represents a clinical assessment of an allergy or intolerance and can be distinguished in Allergy when there is a propensity for hypersensitivity reaction(s) to a substance and Intolerance when the condition is due to a propensity for adverse reactions to a substance that is not judged to be allergic or "allergy-like". The specific substance to which the individual is allergic or intolerant is represented by the “coded” `MedicationOrSubstance` class while the enumeration `CategoryAllergyOrIntolerance` allows to select the category to which the allergy or intolerance belongs to (food, medication, environment and biologic).

Workout model

`Workout`, subclass of `Procedure`, denotes a practice or exercise to test or improve one's fitness, ability, or performance of a person or group of persons. This information is generally gathered by a wearable device, e.g. bracelets fitness tracker.

Demographic model

`DemographicMeasure` represents the measure of a characteristic of a population at a certain time, including demographic, economic and other administrative information. `DemographicMeasure` is specialized in several subclasses:

- `AgeingIndex`: ratio between older (65 years or more) and younger (0 to 14 years) inhabitants, multiplied by 100;
- `EmploymentRate`: percentage of employed inhabitants*LabourMigrationIndex*: ratio between the number of labour-active inhabitants in the territorial unit of working place and the number of labour-active inhabitants in the territorial unit of place of residence, multiplied by 1000;
- `PopulationDensityPerKm2`: number of inhabitants per square km;
- `UnEmploymentShareAmongActiveInhabitants`: share of unemployed among active inhabitants*;*
- `MarriagesNumber`: number of marriages per 1000 inhabitants;
- `NumberOfFinishedApartments`: number of finished apartments per 1000 inhabitants;
- `MigrationFromAbroad`: percentage of migrants of abroad among inhabitants;
- `NumberOfBirths`: number of births per 1000 inhabitants;
- `AverageMonthlyGrossEarnings`: average monthly gross income index (SI=100);
- `KindergartenChildren`: children in kindergartens (as percentage of all children aged 1-5);
- `MunicipalDevelopmentIndex`: aggregated index of municipal development in comparison to national development (SI = 100);
- `HelpAtHome`: age-standardised proportion of help at home programme users (aged 65+ years, per 1000 inhabitants);
- `AdultsMyocardialInfarctionIncidenceRatio`: age-standardised rate of hospital admissions for myocardial infarction in adults (35-75 years per 1000 inhabitants);
- `CancerIncidenceRatio`: age-standardised rate of cancer incidence (as ratio of number of inhabitants);
- `PrimaryLevelEducatedAdults`: number primary-level or less educated per 1000 inhabitants.
- `AntihypertensiveMedicationRatio`: age-standardised percentage of population taking prescription drugs for raised blood pressure in a specific year;
- `HipFracturesRatio`: age-standardised rate of hospital admissions for hip fracture in the elderly (aged 65+ years, as ratio of number of inhabitants);
- `AbsteentismRatio`*,* sick leave days per worker per year*;*
- `AntidiabeticMedicationRatio`: *age-standardised percentage of population taking prescription drugs for diabetes in a specific year;*
- `StrokeIncidenceAdultsRatio`: age-standardised rate of hospital admissions for stroke in adults 35-85 years (as ratio of number of inhabitants);
- `AllCauseMortalityRatio`: age-standardised rate of all-cause mortality (as ratio of number of inhabitants);
- `ChildrenAsthmaRatio`: age-standardised rate of hospital admissions for asthma in children 0-19 years (as ratio of number of inhabitants);
- `MedicationForMentalDisorders`: age-standardised percentage of population taking prescription drugs for raised blood pressure in a specific year.