BioMedBridges

Deliverable 8.2

Project Title:	Building data bridges between biological and medical infrastructures in Europe
Project Acronym:	BioMedBridges
Grant agreement no.:	284209
	Research Infrastructures, FP7 Capacities Specific Programme; [INFRA-2011-2.3.2.] “Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Life sciences"
Deliverable title:	Definition of personalised medicine data types
WP No.	8
Lead Beneficiary:	1: EMBL
WP Title	Use case: Personalized Medicine
Contractual delivery date:	30 June 2014
Actual delivery date:	30 June 2014
WP leader:	16: UH, Imre Vastrik
Contributing partner(s):	1: EMBL, 3: KI, 5: UDUS, 7: TUM-MED, 16: UH

Authors: Imre Vastrik, Helen Parkinson, Julie McMurry, Adam Faulconbridge, Nathalie Conte, Roxana Merino Martinez, Morris Swertz, Bart Charbon

Contents

1 Executive Summary

The purpose of deliverable 8.2 is to describe the data types pertinent for personalised medicine and the links between them. To do this, we first identified a concrete use case: acute myeloid leukemia and then identified the needs of the key stakeholders: physicians, clinical researchers, biomedical researchers, and patients. Guided by these needs we built a prototype data model for structuring, organising and annotating the patient data accordingly. The data model will form the basis for the prototype implementation of an IT solution for storing, management and querying of the personalised medicine data which will then be used to implement and demonstrate the personalised medicine informatics pipeline (deliverable 8.3).

This report starts by briefly explaining the opportunities and challenges related to practising personalised medicine. We then describe a concrete case of applying personalised medicine, analyse the stakeholder needs, enumerate the types of data (both private and public) to actually gain insights into the patient data. The report concludes by specifying a prototype data model for structuring and organising the data in order to meet stakeholder needs.

2 Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	Develop a process for secure sharing of and access to PM data		x
2	Define types of PM data and mapping between them	x	
3	Develop a PM informatics pipeline		x

3 Detailed report on the deliverable

3.1 Background

Personalised medicine is an emerging practice of medicine that uses the genetic profile and other molecular characteristics from a diseased individual or his/her samples in order to guide treatment (Topol 2014). “Omic” technologies, such as “next generation” sequencing, are widely used in research and are also making inroads into the medical practice. As part of systematic clinical studies, as well as in compassionate setting where physician has exhausted other avenues, physicians treating cancer patients have the patients’ tumor genome analysed with the aim of finding drivers of the disease and actionable mutations.

However, physicians do not often have the time or expertise to wade through the raw “omics” analysis; delivering the output concisely in a way they can understand has proven to be challenging. Physicians need summary reports which allow them to drill into the underlying data as and when needed.

Generation of the reports entails integration with the existing knowledge and information. This knowledge is not static but evolves and accumulates over time. As a result the same patient-derived data may provide additional or even different insights at different points in time. Also the recipes to interpret the data may change over time.

Yet another way of gaining insights on the patient from his/her data is by looking at the outcome of other patients with similar molecular characteristics, i.e. case matching. This necessitates that other patients’ data be interoperable, for example using the same data model and/or using the common code sets/ontologies/controlled vocabularies, as well as accessible and “queryable”.

In addition to being useful for diagnosis of the disease and guiding the treatment of the patient, personalised medicine data are also very valuable for research: mining the data for novel associations and generating new hypotheses, testing the validity of new associations and hypotheses, checking the protocol feasibility and identifying patients suitable for clinical trials of targeted drugs resulting in benefits for future patients.

One of the problems hampering the broader use of the personalised medicine data is the lack of suitable and interoperable IT solutions for structuring, organising, managing, using, and querying/analysing data. A secure framework for sharing this data is also of paramount importance and addressed separately in the design presented in deliverable 8.1. Whereas IT-literate users may be happy with downloading or querying large volumes of data which they can interrogate programmatically, the "ordinary" users—physicians, biologists, geneticists etc.—appreciate the possibility to interact with the data via graphical user interface. In addition ethical constraints may preclude bulk downloads. Furthermore, even the computationally advanced users still face the challenge of having to understand how different types of data relate to each other. For example, whereas there are "low level" standards for certain kinds of data, e.g. VCF for describing the genetic variants identified, there is no formal unified/standard/widely accepted agreement/formalism for indicating what was the sample (or sample pair) from which the variants were detected, what tissue and disease state did the sample represent, when was it taken, from whom and what is the medical history of the person from whom the sample(s) originated. This is an active area of development by the Global Alliance for Genomics and Health (GA4GH) and BioMedBridges is represented in the Meta data working group in this activity (co-chair Parkinson). There are several drawbacks to the lack of standards in this area. It hampers data exchange and aggregation of data from multiple sources. It also makes interrogation of external data more difficult since in order to formulate your query correctly, you first need to understand how this dataset is structured.

In this report we list the data types pertinent for personalised medicine of cancer - the use case to which we apply our methodology - and specify the links between those data types. In doing so we have specified the prototype data model for structuring, organising and annotating (by linking to the external knowledge- and reference databases) the patient data pertinent for personalised medicine for our cancer use case. The data model will form the basis for the prototype implementation of an IT solution for storing, management and querying of the personalised medicine data which will then be used to implement and demonstrate the personalised medicine informatics

pipeline (D8.3). This will serve various stakeholders including physicians, clinical and basic researchers as well as patients.

3.2 Concrete use case: story of a leukemia patient

One of the disease areas that is particularly thought to benefit from personalised treatment is cancer. Owing to its etiology (accumulation of mutations in a person's genome combined with environmental effects), cancer is a very heterogeneous disease. Anatomically and histologically-similar tumors can harbour different de-regulated molecular mechanisms resulting in radically different susceptibility to a particular drug. On the other hand, tumors at different anatomical locations can harbour similar mutations and hence similar deregulated molecular mechanisms. This means that drugs which have originally been approved for a certain anatomically-defined cancer could also be used on cancers at different anatomical locations even though the drug may not have been formally demonstrated to have an effect on this particular (anatomically defined) cancer. Whereas such an approach may not be possible in the cases where there is a standard-of-care guideline to follow, sometimes there are no guidelines. This is typically the case with relapsed tumors, where need for customised approaches is particularly acute.

In WP8, we facilitate PM insights through developing data models, tools and bridges. We have based our work on a very concrete use case – the hematologists at the Helsinki University Central Hospital Oncology Clinic treating relapsed acute myeloid leukemia (AML) patients. It is a good test case for personalised medicine since the samples can be taken relatively easily and even in a serial manner, patients are concentrated into a medical speciality, and complex drug treatment regimes are required.

The aliquots of the samples are biobanked at the storage facility affiliated with the local BBMRI node. The samples are subjected to molecular profiling—identification of somatic mutations, estimates of gene copy number changes and gene expression levels—at the local EATRIS infrastructure and analysed for drug sensitivity and resistance at the local EU-OPENSCREEN facility. Equipped with the results of these tests, the hematologists can then make more educated decisions on how to treat a particular patient, for example prediction of response to a particular drug treatment regime.

A typical example of what can be done has been recently published in a Cancer Discovery article by Pemovska et al. 2013. Briefly, the hematologists encountered a 54-year-old AML patient who had failed three consecutive induction therapies. The disease could only be kept in check by chemotherapy which killed also the normal blood cells. As the result the patients was severely immunocompromised and had to be kept in isolation to prevent infection. With essentially no other options available the hematologists subjected the bone marrow sample to drug sensitivity and resistance testing (DSRT). The DSRT results highlighted dasatinib, sunitinib, and temsirolimus among the top five most selective approved drugs. In an off-label compassionate use setting, the patient received a combination of these targeted drugs, resulting in rapid reduction of the bone marrow blast count and marked improvement in the poor performance status. Concomitantly, the blood counts rapidly normalized resulting in complete remission with incomplete platelet recovery. The blast cell count, indicative of the disease burden, decreased. The neutrophil count, indicative of the functioning immune system, increased. The patient's condition improved, albeit nevertheless temporarily, so markedly that he was able to go home after spending half a year at the isolation ward. This provides a working example of how personalised medicine can contribute directly to medical treatment.

3.3 Stakeholder analysis

In order to understand which data types should be included in the data model and how they should be structured, we identified the key stakeholders and analysed their respective needs. One key difference between physicians and researchers is that the physician is focused on diverse kinds of data from a single patient; the physician may want to apply any existing general knowledge to interpret a specific patient's data and make an informed treatment choice. By contrast, researchers typically analyse certain slices of data aggregated over many patients in order to contribute to general knowledge. Table 1 explores these stakeholder needs in more depth. The external resources useful for addressing the needs are indicated in the Appendix A. As addressed in deliverable D8.1, these diverse data usage patterns also require different modes of secure access control.

Table 1. Stakeholders with needs for personalised medicine data

Stakeholder	What the stakeholder wants and needs from the data
Physician	• List of mutated genes (including point mutations, copy number variation and fusions) likely to drive the disease and highlighting the ones for which specific drugs are available. • Shortlist of the drugs (including those off-label) that according to the ex vivo drug sensitivity testing data could be beneficial for treating the disease. • Pointers to the clinical trials that are recruiting at the time and which are appropriate for the patient either according to the diagnosis, mutated genes, or ex vivo drug sensitivity. • Predict the outcome for this patient based on the previous patients with similar molecular profile. (Eg. is my gene X in a region susceptible to copy number aberration in other cancers)?
Clinical researcher	• Assess feasibility of a clinical trial protocol, i.e. estimate the number of patients that (will) match the criteria to be included in a clinical trial. • Identify a cohort of patients that match the criteria to be included in a clinical trial.
Basic biomedical researcher	• Identify a collection of data matching certain criteria for further in-depth analysis, e.g. mining for new associations or trying to replicate previous finding in a new dataset. • Identify samples from patients with a disease of interest to be subjected to molecular profiling. • Identification of new variants with related phenotypic information and comparison of these with existing data e.g. Cancer Genome Atlas (TCGA) or COSMIC

3.4 Types of data relevant in personalised medicine

The personalised medicine data entails both the “traditional” medical record data as well as the “omics” data that has thus far been considered to belong to the realm of research data. Conceptually the omics data is similar to clinical chemistry laboratory measurements which measure (or estimate) the existence or amount of some analyte (e.g. mutation, gene expression level, protein phosphorylation state) from or characteristic (sensitivity to a drug ex vivo) of the patient sample. Omics technologies deploy a broad battery of

analytes or characteristics measured simultaneously instead of using analyte or characteristic individual assays as in clinical blood chemistry.

Table 2. Data types, standards, and integration paths to address the needs of stakeholders.

Data type	Example	Relevant standards	Integration paths
Condition: Health condition of the person, i.e. the diagnosis. ICD-10 codes are primarily used with ICD-O-3 codes enabling finer grained classification if and when this becomes available (i.e. as the result of diagnostic test being performed).	C91.0 Acute lymphatic leukemia	ICD-10, ICD-O-3, SNOMED-CT, NCIT	Interoperability between ontologies e.g. Disease Ontology, SnoMed-CT, Disease Ontology - via cross references and curation of these.
Medication: The medicine given to the person together with dosing and timing information. Currently the drugs are identified by just their compound names and are not explicitly backed by any particular codeset.	Medication: Cytarabine; Daily dose: 100.0; Unit: mg/m2	Anatomical Therapeutic Chemical (ATC) classification system could be one source of standardised terminology that would cover all the approved drugs. However, as this would not cover the investigational drugs possibly given during clinical trials another codeset, such as ChEMBL or DrugBank, would be better.	ChEMBL, DrugBank
Disease status: Outcome of / response to the treatment. Similar to diagnosis but more fine grained. Uses bespoke codeset for each disease (e.g. AML, CMML etc) and also specifies which code to use in a given situation. For example, the code CR-MRDneg stands for “Remission, minimal residual disease negative”, meaning that the patient is morphologically in a remission, is known to have markers for detecting minimal residual disease with these nevertheless being undetectable.	CR-MRDneg	The response categories are based on recommendations from international expert panels. For example, the AML response criteria and survival outcomes are described in Döhner et al. 2010. However, to our best knowledge as yet these have not been turned into a formal standard represented in a structured manner.	Codify these and share publicly.
Samples: The samples taken from the patient for biobanking and omics analysis.	Mononuclear cells	MIABIS, ICD-O-3 topography section, SPREC codes	Inclusion in the BioSamples database if consent permits. Consistent local identifier and standards use if consent does not permit
Observation: Observations and measurements performed on the person as well as results of diagnostics tests.. These could be height, weight, blood pressure, estimates of the size of certain organs, e.g. spleen. As the things stand the terminology used to	Observable: Spleen size as measured with ultrasound; Value: 14.1; Unit: cm	Logical Observation Identifiers Names and Codes (LOINC), a database and universal standard for identifying medical laboratory observations and measurements, could be a source of standardised terminology to provide	

specify the observation (e.g. height) is currently not coming from any ontology, controlled vocabulary or codeset.	Observable: Blood hemoglobin Value: 171 Unit: g/l Observable: Bone marrow blasts Value: 76 Unit: %	interoperability with same data collected elsewhere. Finland has a national codeset for laboratory tests which is used nationwide in the healthcare organisations to identify the diagnostic tests. However, the terminology in the codeset is just in the official languages of Finland (Finnish and Swedish) and no mapping to the internationally more widely used LOINC standard is readily provided.	
Omics analysis results: Somatic mutations - specifically single nucleotide variants (SNV) and small insertions and deletions (indels) - and their frequency in the sample analysed. The location of the variant is indicated in the chromosomal coordinates, the sequence change is indicated by both the reference sequence at the given location and mutated sequence found in the sample. The frequency is calculated based on the number of reads supporting the reference sequence and those supporting the variant.	Chromosome: chr10 Position: 8111432 Reference: TCA Variant: T Frequency: 23.42%	The Locus Reference Genomic (LRG; http://www.lrg-sequence.org/; MacArthur et al., 2014; Dalgleish et al., 2010) Human Genome Variation Society (HGVS; http://www.hgvs.org/)	Variant Effect Predictor, Cosmic
Omics analysis results: Gene copy number status in the sample analysed.	Gene: ENSG00000077782 Copy number status: gain	Genes are identified by their Ensembl gene models and the copy number status is indicated by labels such as “diploid”, “gain”, “heterozygous_deletion” and “homozygous_deletion”.	
Omics analysis results: Gene expression levels in the tumor tissue. Expression levels are given in Fragments Per Kilobase of exon per Million fragments mapped (FPKM)	Gene: ENSG00000170345 Value: 303.372 Unit: FPKM	Genes are identified by their Ensembl gene models.	
Omics analysis results: Fusion genes found on the basis of expressed fusion transcripts in the tumor tissue.	5’ gene: ENSG00000110713 3’ gene: ENSG00000165671	Genes are identified by their Ensembl gene models.	
Omics analysis results: Sensitivity of the tumor tissue to compounds (including drugs) as tested ex vivo. Cell viability is used as a read-out. Each compound is tested at a range of different concentrations allowing to derive a dose response curve. The dose response curve of a given drug for the tumor cells is compared with that for the normal/healthy control cells. The difference between the areas under the dose response curves is denoted as selective drug sensitivity score (sDSS, Yadav et al., 2014) and use as a measure of the drug’s ability to decrease cell viability (by inhibition of growth or induction of cell death).	Compound: Sunitinib sDSS: 16.1	Anatomical Therapeutic Chemical (ATC) classification system could be one source of standardised terminology that would cover all the approved drugs. However, as this would not cover the investigational drugs possibly given during clinical trials another codeset, such as ChEMBL or DrugBank, would be better.	

3.6 Personalised Medicine data model

Once the list of necessary data elements (table 2) was established, it was important to formally structure the relationships between these elements to support the stakeholders' questions. Figure 1 shows a summary view of the resulting data model. The model is loosely based on the data model behind the Finnish Health Account Taltioni (http://taltioni.fi/en/). At the center of the model is Person, i.e. patient, who may be:

- **diagnosed with medical conditions** represented as instances of class `Condition`. Every instance records the start and end time of the condition, time of the diagnosis, person making the diagnosis as well as the diagnosis code. Diagnosis codes are represented as instances of the class `Disease` and correspond to the entries in the ICD-10 and ICD-O-3 code sets.
- **taking/given medications**. An instance of the `Medication` class captures the timing of the treatment and the active ingredient of the drug given/taken. The latter is represented as an instance of class `Chemical` and corresponds to an entry in the ChEMBL (or DrugBank) database.
- **subjected to various observations** such as clinical laboratory tests, measurements and questions. An instance of the class `Observation` captures the timing, observed value and the observable. The latter is represented as an instance of class `Observable` and corresponds to an entry in the LOINC database.
- **providing samples** for biobanking and advanced laboratory analyses such as those identifying somatics mutations, gene copy number changes and sensitivity to drugs. An instance of the class `TissueSample` captures sampling time, link to the storage facility IT system and type of the tissue. The latter is represented as as an instance of class `Tissue` and corresponds to an entry in a tissue controlled vocabulary/ontology such as the ICD-0-3 topography section.

  The sample logistics related information could also in principle be

included in the data model. However, we nevertheless felt that considering the use case this would have been a step too far.

The samples may be subjected to various omics analysis runs each of which is represented as an instance of an appropriate `Analysis` sub-class (see Appendix B for the class hierarchy). This instance captures the time of the analysis, protocol used and points to a set of analysis results of which there are typically many since the same omics analysis measures multiple "analytes" (e.g. genes, variants or drug sensitivities) in one go. Each analyte measurement is represented as an instance of an appropriate `AnalysisResultValue` sub-class, i.e. `SomaticMutation`, `GeneCNV`, `DrugSensitivity`, instance.

For example, the results of the gene copy number variation (CNV) analysis are captured as instances of the `GeneCNV` class with one instance for each gene assessed. The `GeneCNV` instance captures the gene being assessed and its copy number status. A gene is represented as an instance of class `Gene` and corresponds to a gene entry in the Ensembl database. Also the somatic mutation analysis results utilise the `Gene` instances (albeit via an instance of class `ReferenceMutation`) to indicate the gene containing the variant.

Similarly, the results of the drug sensitivity analysis are captured as instances of the `DrugSensitivity` class with one instance for each compound tested. This instance captures the effect of the compound in the form of selective drug sensitivity score (sDSS), intermediate data items used to assess this (EC50, dose reponse curve slope etc) as well as the compound tested. The latter is represented as an instance of class `Chemical` which is also used to describe the medication the person may be having. Such a "sharing" of instance provides a way of internal integration of different types of data.

The same design pattern can also be used to extend the data model to cover other omics data types.

Another key principle of data organisation is that whenever possible existing code sets, vocabularies, ontologies and databases are used in order to provide interoperability with the external information (Appendix A). Briefly, the

internal representation of diagnosis instances corresponds to ICD entries, compounds (including drugs) to ChEMBL (or DrugBank) entries, measurements, tests and observables to LOINC entries, genes to Ensembl gene entries. The concrete benefits are:

- Facilitates discoverability and linking in. The user coming from outside can find and interrogate the data source using “standard terminology”.
- Facilitates linking out as well as dynamic retrieval of additional information from external resources, such as mutation frequency of a gene in other cancers, if a gene is known to be a tumor driver gene or suppressor, drugs targeting a gene (product), clinical trials with a specific drug, clinical trials for a specific disease, tissue samples representing the same diagnosis, pathways where a gene/protein is involved, drugs targeting the pathways containing a specific gene etc.
- Helps to enforce data integrity. With appropriate tooling the user won’t be able to use any other terminology than the one specified by the data mode (or rather the data itself).

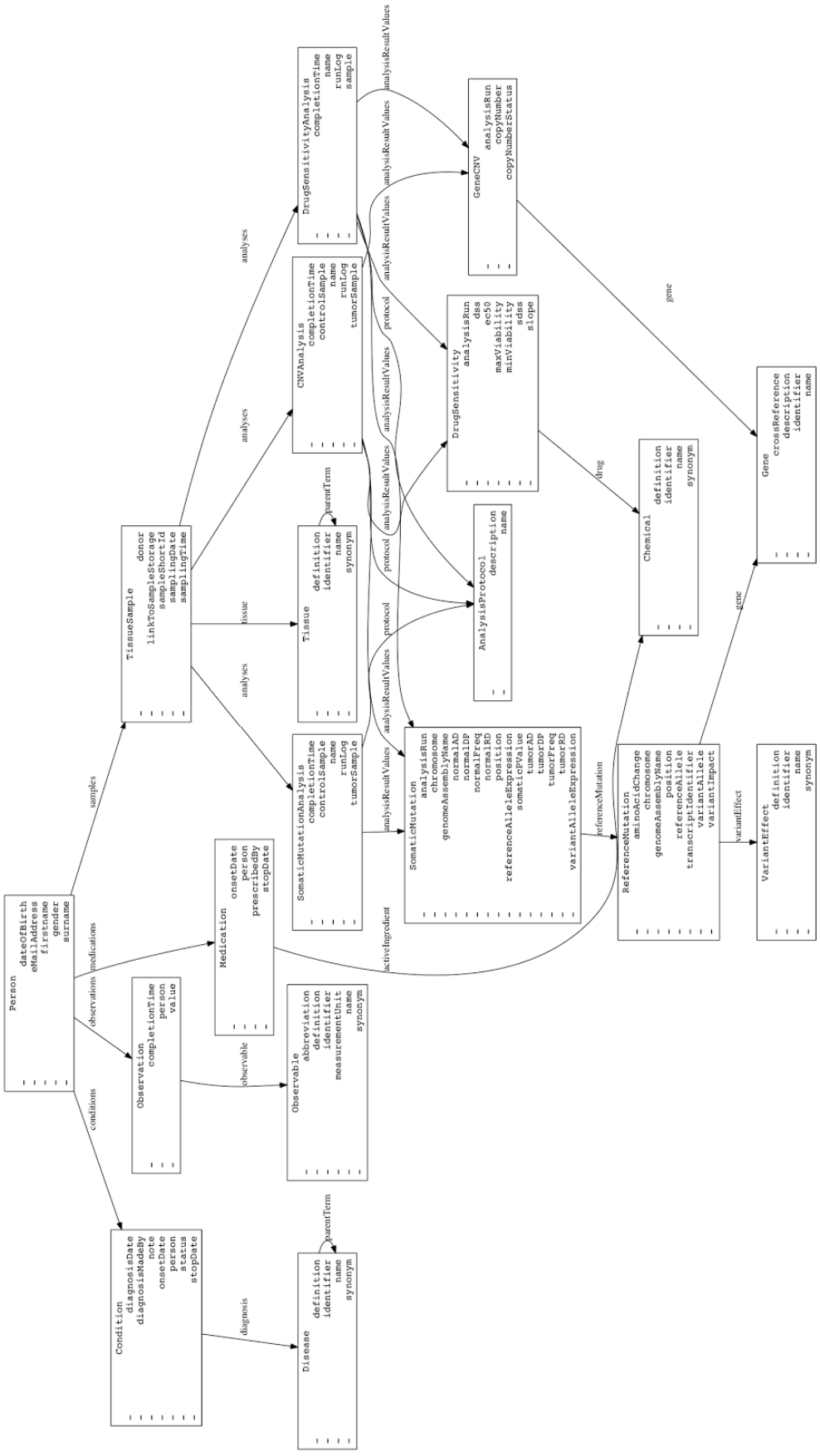

Figure 1. Simplified view of the personalised medicine data model

3.7 Future Work

- Map WP8 data types outcomes to WP3 activities on identifiers best practice (D3.1).
- Implement data types and data model in D8.3.
- Use the PM use case to drive semantic integration informing D3.4 by testing semantic interoperability with data derived from this work package.
- Define a set of PM standards to be added to the standards registry D3.2 and implement these in future versions of the infrastructure described here.

4 List of appendices

Appendix A Cancer resource integration - catalog of use cases, data types and external resources for integration

Appendix B Prototype personalised medicine data model - a complete representation of the data model

5 Background Information

This deliverable relates to WP8 Use case: Personalized Medicine. Background information on this WP as originally indicated in the description of work (DoW) is included below.

WP8 Title: Use case: Personalized Medicine
Lead: 16: UH
Participants: EMBL, KI, UDUS, TUM-MED, UH

WP8 will integrate complex data sets to understand disease pathogenesis and improve biomarker and treatment selection

Work package number	WP 8	**Start date or starting event:**	month 1
Work package title	Use case: Personalized Medicine		
Activity Type	RTD		

Participant number	1: EMBL	3: KI	5: UDUS	7: TUM-MED	16: UH			
Person-months per participant	16	8	5	8	32			

Objectives

1) Definition of a process for secure sharing of and access to personalized medicine (PM) data.
2) Definition of existing PM data types and mappings between them.
3) Pilot the use of PM data to support the clinical decision making process.

Description of work and role of participants

Use case: Personalized Medicine - integrating complex data sets to understand disease pathogenesis and improve biomarker and treatment selection

Task 1. Develop a process for secure sharing of and access to PM data

Building on the work carried out in Secure access work package (WP5) we will develop a process by which a producer of the data can share and the user of the data can gain access to the PM data in a secure, legal yet easiest possible manner. FIMM will have the role of a prototype PM data provider as well as a user. TUM, as the leader of WP5 will provide expertise in privacy protection as well as secure sharing and access matters.

Task 2. Define types of PM data and mapping between them
Measurements made with different technologies may not be (and usually are not) directly comparable even though the underlying thing measured (e.g. certain mRNA level) may be exactly the same. This creates a situation where a user of the data may inadvertently be "comparing apples with oranges". To avoid that we will catalogue data types (as well as pertinent standards) relevant to PM and provide mapping between them if applicable. FIMM will provide PM domain expertise. KI, as the leader of the Standards work package (WP3) will provide know-how of existing standards.

Task 3. Develop a PM informatics pipeline
As a proof of concept that the tasks above facilitate the interoperability of different PM data types we will develop a prototype PM informatics pipeline to support the decision making process in PM. This prototype pipeline will utilise the data type specifications and standards established in Task 2 and be subject to constraints of access procedures established in Task 1. FIMM will be a prototype PM data producer and user. EMBL-EBI as the leader of Technical Integration work package (WP4) will provide expertise on general framework and architecture of the implementation.

Deliverables

No.	Name	Due month
D8.1	Process specification for secure sharing of and access to PM data	30
D8.2	Definition of PM data types (report)	30
D8.3	Demonstration of interoperability between different types of PM data	48

6 References

Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, Chen Y, McLaren WM, Larsson P, Vaughan BW, Béroud C, Dobson G, Lehväslaiho H, Taschner PE, den Dunnen JT, Devereau A, Birney E, Brookes AJ, Maglott DR. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med. 2010 Apr 15;2(4):24. doi: 10.1186/gm145. PubMed PMID: 20398331; PubMed Central PMCID: PMC2873802.

Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz MA, Sierra J, Tallman MS, Löwenberg B, Bloomfield CD; European LeukemiaNet. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010 Jan 21;115(3):453-74. doi: 10.1182/blood-2009-07-235358. Epub 2009 Oct 30. Review. PubMed PMID: 19880497.

MacArthur JA, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, Larsson P, Flicek P, Dalgleish R, Maglott DR, Cunningham F. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic Acids Res. 2014 Jan;42(Database issue):D873-8. doi: 10.1093/nar/gkt1198. Epub 2013 Nov 26. PubMed PMID: 24285302; PubMed Central PMCID: PMC3965024.

Pemovska T, Kontro M, Yadav B, Edgren H, Eldfors S, Szwajda A, Almusa H, Bespalov MM, Ellonen P, Elonen E, Gjertsen BT, Karjalainen R, Kulesskiy E, Lagström S, Lehto A, Lepistö M, Lundán T, Majumder MM, Marti JM, Mattila P, Murumägi A, Mustjoki S, Palva A, Parsons A, Pirttinen T, Rämet ME, Suvela M, Turunen L, Västrik I, Wolf M, Knowles J, Aittokallio T, Heckman CA, Porkka K, Kallioniemi O, Wennerberg K. Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. Cancer Discov. 2013 Dec;3(12):1416-29. doi: 10.1158/2159-8290.CD-13-0350. Epub 2013 Sep 20. PubMed PMID: 24056683.

Topol EJ. Individualized medicine from prewomb to tomb. Cell. 2014 Mar 27;157(1):241-53. doi: 10.1016/j.cell.2014.02.012. Review. PubMed PMID: 24679539; PubMed Central PMCID: PMC3995127.

Yadav B, Pemovska T, Szwajda A, Kulesskiy E, Kontro M, Karjalainen R, Majumder MM, Malani D, Murumägi A, Knowles J, Porkka K, Heckman C, Kallioniemi O, Wennerberg K, Aittokallio T. Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. Sci Rep. 2014 Jun 5;4:5193. doi: 10.1038/srep05193. PubMed PMID: 24898935; PubMed Central PMCID: PMC4046135.

Appendix A: Cancer resource integration

Stakeholder	Question	External source	Data type	Program-matic access
Basic biomedical researcher	Where can I find raw data of experiments with my disease X?	ArrayExpress	database of functional genomics experiments	Yes
Basic biomedical researcher	Where can I find more samples with my disease X?	BBMRI database metadata, BioSamples	database of BBMRI related metadata	No
Basic biomedical researcher	Where can I find more samples with my disease X?	Biosamples Database	database of biosamples	Yes
Basic biomedical researcher	Where can I find relevant animal model information?	caMOD	cancer model database	No
Physician	Does the protein target have a corresponding known drug? If not, might it be druggable?	canSAR	An integrated database that brings together biological, chemical, pharmacological (and eventually clinical) data.	No
Physician	Is my gene X in a region suceptible to copy number abberation in other cancer?	CGHviewer	copy number and genotyping cancer cell line database	No
Physician	Does the protein target have a corresponding known drug? If not, might it be druggable?	ChEBI	drug database	No
Physician	Does the protein target have a corresponding known drug? If not, might it be druggable?	ChEMBL	drugs database	Yes
Clinical researcher	Is there a clinical trial currently recruiting subjects with this disease?	Clinical Trials Information Mediator (ECRIN)	Semi-structured trials metadata	Yes
Clinical researcher	Is there a clinical trial currently recruiting subjects with this disease?	Clinical Trials.gov	Semi-structured trials metadata	No
Basic biomedical researcher	Where can I find more cases carrying my mutation in gene X in a disease context (the same of a different one)?	COSMIC	human cancer literature database	Yes
Basic biomedical researcher	Where can I find processed data of experiments where my gene of interest is mutated ?	COSMIC	human cancer literature database	Yes
Physician	Does the protein target have a corresponding known drug? If not, might it be druggable?	DrugBank	drug database	No
Basic biomedical researcher	Is there a relevant cell line drug screen?	DTP Human Tumor Cell Line Screen	drug screening database	No
Basic biomedical researcher	Where can I find raw data of experiments with my disease X?	EGA	Genome-phenome database	Yes
Basic biomedical researcher	Where can I find relevant animal model information?	eMICE	animal model database	No
Basic biomedical researcher	What is my gene/transcript annotation	ENSEMBL	genome databases	Yes
Clinical researcher	Is there a clinical trial currently recruiting subjects with this disease?	EudraCT	Semi-structured trials metadata	No
Physician	What is the drug resitance/sensitivity from tumour ex vivo culture?	ex-vivo drug sensitivity and resistance testing (DSRT)	drug screening ASSAY	No

Stakeholder	Question	External source	Data type	Program-matic access
Basic biomedical researcher	Where can I find processed data of experiments where my gene of interest is mutated ?	Gene expression Atlas	database of functional genomics experiments	No
Basic biomedical researcher	Where can I find more cases carrying my mutation in gene X in a disease context (the same of a different one)?	ICGC data portal	database of ICGC	Yes
Basic biomedical researcher	Where can I find more samples with my disease X?	ICGC data portal	database of ICGC	No
Basic biomedical researcher	Where can I find processed data of experiments where my gene of interest is mutated ?	ICGC data portal	database of ICGC	Yes
Physician	What can we learn from multiple omics data on the same sample?	inside institute	tab format?	No
Basic biomedical researcher	What are the interaction partners of my protein?	IntAct	protein database	Yes
Physician	Is drug repositioning (off-label use) a possibility?	MANTRA	computational tool	No
Basic biomedical researcher	Where can I find other relevant metabolomics experiments	Metabolights	database for Metabolomics experiments and derived information	No
Basic biomedical researcher	What is my protein structure?	PDB	protein database	No
Basic biomedical researcher	What are the pathways were my protein is involved?	Reactome	protein database	Yes
Basic biomedical researcher	What is my protein annotation?	UniProt	protein database	Yes
Physician	What is the patient sample annotation?		metadata (age, environemental data)	No
Physician	What is the patient's tumour diagnosis?		pathology report	No

Appendix B. Prototype personalised medicine data model

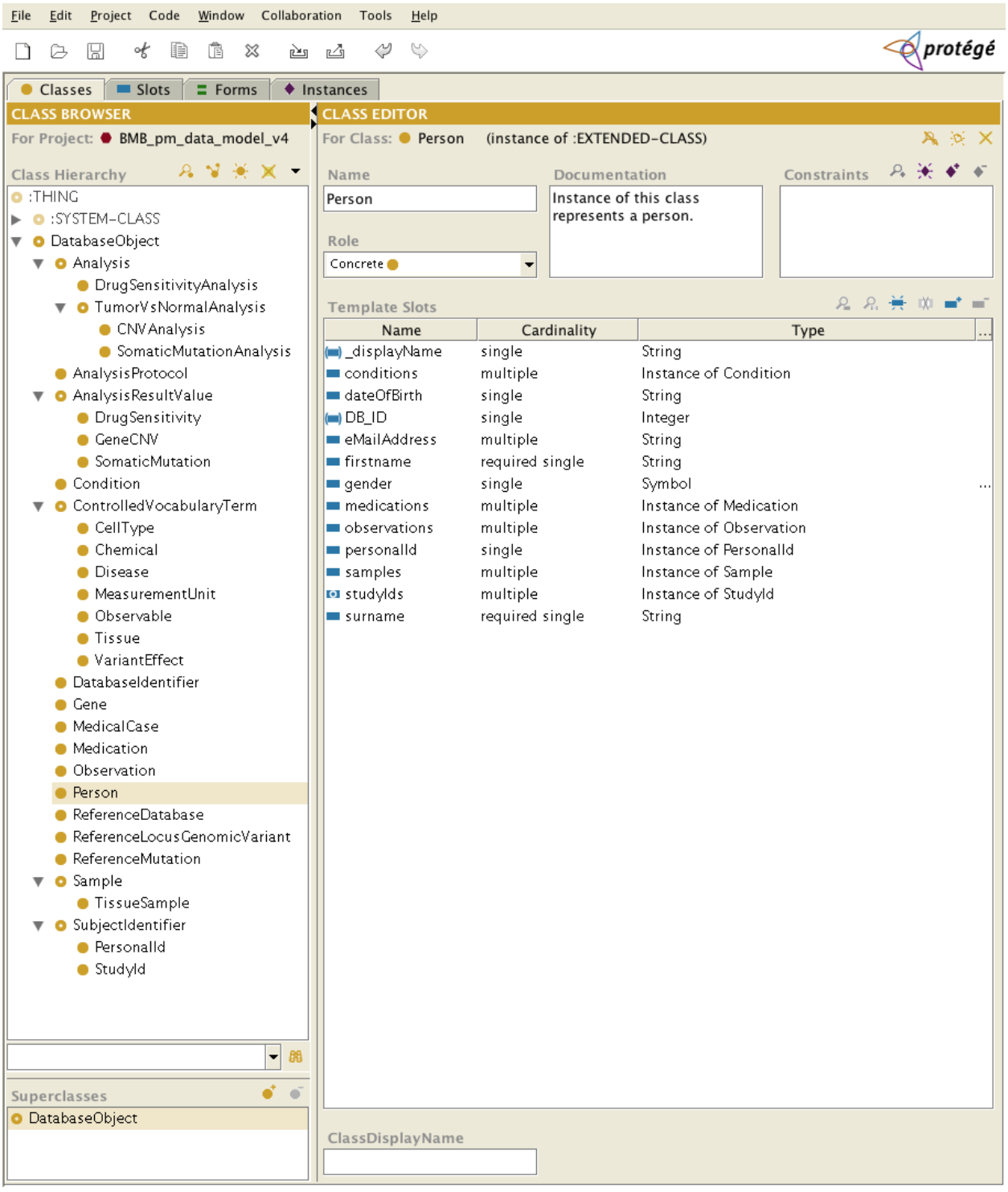

Figure 1. Screen-shot from the Protege tool used for data model specification. Class hierarchy is shown in the left-hand pane.

Data model classes

Class 'DatabaseObject'

An abstract superclass of all classes.

superclass:				
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin

Class 'ControlledVocabularyTerm'

An abstract superclass for external controlled vocabularies each of which will be prepresented as a subclass. The terms are represented as instances of the subclass.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'Condition'

Instance of this class describes a condition e.g. a diagnosis of a person at specified time.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
diagnosis	1	INSTANCE	Disease	Condition
diagnosisDate	1	DATE	N/A	Condition
diagnosisMadeBy	1	INSTANCE	Person	Condition
note	1	TEXT	N/A	Condition
onsetDate	1	DATE	N/A	Condition
person	1	INSTANCE	Person	Condition
status	1	TEXT	N/A	Condition
stopDate	1	DATE	N/A	Condition

Class 'Gene'

Instance of this class represents a gene record in an external database of genes. As such this class could have been put as a sub-class of ControlledVocabularyTerm

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
cancerGeneCensus	1	INSTANCE	DatabaseIdentifier	Gene
crossReference	+	INSTANCE	DatabaseIdentifier	Gene
description	1	TEXT	N/A	Gene
identifier	1	TEXT	N/A	Gene

name	1	TEXT	N/A	Gene
referenceDatabase	1	INSTANCE	ReferenceDatabase	Gene

Class 'MedicalCase'

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analyses	+	INSTANCE	Analysis	MedicalCase
condition	1	INSTANCE	Condition	MedicalCase
patient	1	INSTANCE	Person	MedicalCase
samples	+	INSTANCE	Sample	MedicalCase

Class 'AnalysisProtocol'

Instance of this class represent a particular protocol (workflow) version used to perform certain analysis. E.g. a particular workflow for calling CNVs would be one instance. All analysis performed according to this protocol would refer to this instance. If the protocol is changed, this will be stored as another instance pointed by all the analyses performed accrdingly.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
description	1	TEXT	N/A	AnalysisProtocol
name	+	TEXT	N/A	AnalysisProtocol

Class 'ReferenceLocusGenomicVariant'

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
position	1	INTEGER	N/A	ReferenceLocusGenomicVariant
referenceAllele	1	TEXT	N/A	ReferenceLocusGenomicVariant
referenceLocusGenomicId	1	TEXT	N/A	ReferenceLocusGenomicVariant
variantAllele	1	TEXT	N/A	ReferenceLocusGenomicVariant

Class 'Analysis'

Abstract (super)class representing analysis performed on person's samples or data.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisResultValues	+	INSTANCE	AnalysisResultValue	Analysis
completionTime	1	DATETIME	N/A	Analysis
name	+	TEXT	N/A	Analysis
protocol	1	INSTANCE	AnalysisProtocol	Analysis

runLog	1	TEXT	N/A	Analysis

Class 'SubjectIdentifier'

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
identifier	1	TEXT	N/A	SubjectIdentifier

Class 'ReferenceDatabase'

Instance of this class represents an external (reference) database. Essentially this gives the namespace to the DatabaseIdentifier instance it is attached to. This way access URLs etc can be stored as data rather than in code.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
accessUrl	1	TEXT	N/A	ReferenceDatabase
name	+	TEXT	N/A	ReferenceDatabase
url	1	TEXT	N/A	ReferenceDatabase

Class 'Person'

Instance of this class represents a person.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
conditions	+	INSTANCE	Condition	Person
dateOfBirth	1	DATE	N/A	Person
eMailAddress	+	TEXT	N/A	Person
firstname	1	TEXT	N/A	Person
gender	1	TEXT	N/A	Person
medications	+	INSTANCE	Medication	Person
observations	+	INSTANCE	Observation	Person
personalId	1	INSTANCE	PersonalId	Person
samples	+	INSTANCE	Sample	Person
studyIds	+	INSTANCE	StudyId	Person
surname	1	TEXT	N/A	Person

Class 'Sample'

An abstract superclass to "pull together" all samples. Another sub-class could be StoolSample or perhaps even CeelCultureSample.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analyses	+	INSTANCE	Analysis	Sample

sampleLongId	1	TEXT	N/A	Sample
sampleShortId	1	TEXT	N/A	Sample

Class 'Observation'

Instance of this class represent values of measurements performed and test and observations made on the person or samples drived from it. Examples of this are blood pressure, body temperature, alchohol consumption, various clinical chemistry tests etc.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
completionTime	1	DATETIME	N/A	Observation
observable	1	INSTANCE	Observable	Observation
person	1	INSTANCE	Person	Observation
value	1	FLOAT	N/A	Observation

Class 'AnalysisResultValue'

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisRun	1	INSTANCE	Analysis	AnalysisResultValue
personStudyId	1	TEXT	N/A	AnalysisResultValue
sampleLongId	1	TEXT	N/A	AnalysisResultValue

Class 'DatabaseIdentifier'

Instance of this class represents an identifer in some external database. The database is specified is the instance of the ReferenceDatabase class in the referenceDatabase slot.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
identifier	1	TEXT	N/A	DatabaseIdentifier
referenceDatabase	1	INSTANCE	ReferenceDatabase	DatabaseIdentifier

Class 'ReferenceMutation'

Instances of this class are used to capture the chromosomal position and content of the variant. Used by eg SomaticMutation to store the coordinate and change.\nThis will change as LRGs become more of a reality. Then the chr coordinates can be removed.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
aminoAcidChange	1	TEXT	N/A	ReferenceMutation
chromosome	1	TEXT	N/A	ReferenceMutation
gene	1	INSTANCE	Gene	ReferenceMutation
genomeAssemblyName	1	TEXT	N/A	ReferenceMutation
position	1	INTEGER	N/A	ReferenceMutation

referenceAllele	1	TEXT	N/A	ReferenceMutation
referenceLocusGenomicVariant	1	INSTANCE	ReferenceLocusGenomicVariant	ReferenceMutation
transcriptIdentifier	1	TEXT	N/A	ReferenceMutation
variantAllele	1	TEXT	N/A	ReferenceMutation
variantEffect	1	INSTANCE	VariantEffect	ReferenceMutation
variantImpact	1	TEXT	N/A	ReferenceMutation

Class 'Medication'

Instance of this class describes a medication that the person has/is adminsitered at specified time.

superclass:	DatabaseObject			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
activeIngredient	+	INSTANCE	Chemical	Medication
onsetDate	1	DATE	N/A	Medication
person	1	INSTANCE	Person	Medication
prescribedBy	1	INSTANCE	Person	Medication
stopDate	1	DATE	N/A	Medication

Class 'MeasurementUnit'

Instances of this class represent different units e.g. "nmol/l", "g/l", etc. The idea is that this forces the user to use just one of pre-existing units as opposed to using free text of enum.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'CellType'

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'StudyId'

Instance of this class represents a study-specific identifier. That is generally much less sensitive and can be used by e.g. lab personnel without the need to know or access the social security number.

superclass:	SubjectIdentifier			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
identifier	1	TEXT	N/A	SubjectIdentifier
study	1	INSTANCE	ReferenceDatabase	StudyId

Class 'Tissue'

Instance of this class represents a tissue vocabulary term e.g. a term from ICD-O-3 topography section.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
parentTerm	1	INSTANCE	Tissue	Tissue
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'TumorVsNormalAnalysis'

superclass:	Analysis			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisResultValues	+	INSTANCE	AnalysisResultValue	Analysis
completionTime	1	DATETIME	N/A	Analysis
controlSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis
name	+	TEXT	N/A	Analysis
protocol	1	INSTANCE	AnalysisProtocol	Analysis
runLog	1	TEXT	N/A	Analysis
tumorSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis

Class 'TissueSample'

Instance of this class represent a tissue sample from a person. The type of tissue is specified by the value of the tissue-attribute.

superclass:	Sample			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analyses	+	INSTANCE	Analysis	Sample
donor	1	INSTANCE	Person	TissueSample
donorId	1	INSTANCE	SubjectIdentifier	TissueSample
linkToSampleStorage	1	TEXT	N/A	TissueSample
sampleLongId	1	TEXT	N/A	Sample
sampleShortId	1	TEXT	N/A	Sample
samplingDate	1	DATE	N/A	TissueSample
samplingTime	1	TIME	N/A	TissueSample

tissue	1	INSTANCE	Tissue	TissueSample

Class 'VariantEffect'

Instance of this class specifies the effect the variant has on a transcript, i.e. is it synonomous or non-synonomous, affects slicing, is in 3' UTR etc.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'Observable'

Instance of this class represents lab test, measurement or observation as specified by LOINC or similar codesets.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
abbreviation	1	TEXT	N/A	Observable
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
measurementUnit	1	TEXT	N/A	Observable
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'PersonalId'

Instance of this class represents social security number or similar. Access to this class instances would be fairly restricted.

superclass:	SubjectIdentifier			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
identifier	1	TEXT	N/A	SubjectIdentifier

Class 'Chemical'

Instance of this class represent a chemical entity such as a drug and is equivalent to e.g. ChEMBL (or other suitable/relevant vocabulary of compounds) record.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
fimm_drug_id	1	TEXT	N/A	Chemical

identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'Disease'

Instance of this class represents a diagnosis term and correponds to a ICD-10 and/or ICD-O-3 term/record. The parentTerm attribute is used to capture the term hierarchy.

superclass:	ControlledVocabularyTerm			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
definition	1	TEXT	N/A	ControlledVocabularyTerm
identifier	1	TEXT	N/A	ControlledVocabularyTerm
name	+	TEXT	N/A	ControlledVocabularyTerm
parentTerm	1	INSTANCE	Disease	Disease
referenceDatabase	1	INSTANCE	ReferenceDatabase	ControlledVocabularyTerm
synonym	+	TEXT	N/A	ControlledVocabularyTerm

Class 'DrugSensitivity'

Instance of this class represents a measurement (estimate) of a particular sample's sensitivity to a particular drug. If a sample is tested with 300 drugs, this results in 300 instances etc.

superclass:	AnalysisResultValue			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisRun	1	INSTANCE	DrugSensitivityAnalysis	AnalysisResultValue
drug	1	INSTANCE	Chemical	DrugSensitivity
dss	1	FLOAT	N/A	DrugSensitivity
ec50	1	FLOAT	N/A	DrugSensitivity
maxViability	1	FLOAT	N/A	DrugSensitivity
minViability	1	FLOAT	N/A	DrugSensitivity
personStudyId	1	TEXT	N/A	AnalysisResultValue
sampleLongId	1	TEXT	N/A	AnalysisResultValue
sdss	1	FLOAT	N/A	DrugSensitivity
slope	1	FLOAT	N/A	DrugSensitivity

Class 'DrugSensitivityAnalysis'

Instance of this class represents drug sensitivity analysis (DSRT) of a patient's tumor sample.

superclass:	Analysis			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisResultValues	+	INSTANCE	AnalysisResultValue	Analysis
completionTime	1	DATETIME	N/A	Analysis
name	+	TEXT	N/A	Analysis
protocol	1	INSTANCE	AnalysisProtocol	Analysis
runLog	1	TEXT	N/A	Analysis
sample	1	INSTANCE	TissueSample	DrugSensitivityAnalysis

Class 'SomaticMutation'

Instance of this class represents a call of a somatic mutation (SNV or small indel) as estimated on the basis of teh sequencing data of a tumor-normal sample pair from a patient. This class/instance captures just the analysis-specific info. Position-specific info, e.g. the consequence of the mutation, is put into instances of ReferenceMutation class.

superclass:	AnalysisResultValue			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisRun	1	INSTANCE	SomaticMutationAnalysis	AnalysisResultValue
chromosome	1	TEXT	N/A	SomaticMutation
genomeAssemblyName	1	TEXT	N/A	SomaticMutation
normalAD	1	INTEGER	N/A	SomaticMutation
normalDP	1	INTEGER	N/A	SomaticMutation
normalFreq	1	FLOAT	N/A	SomaticMutation
normalRD	1	INTEGER	N/A	SomaticMutation
personStudyId	1	TEXT	N/A	AnalysisResultValue
position	1	INTEGER	N/A	SomaticMutation
referenceAlleleExpression	1	TEXT	N/A	SomaticMutation
referenceMutation	1	INSTANCE	ReferenceMutation	SomaticMutation
sampleLongId	1	TEXT	N/A	AnalysisResultValue
somaticPValue	1	FLOAT	N/A	SomaticMutation
tumorAD	1	INTEGER	N/A	SomaticMutation
tumorDP	1	INTEGER	N/A	SomaticMutation
tumorFreq	1	FLOAT	N/A	SomaticMutation
tumorRD	1	INTEGER	N/A	SomaticMutation
variantAlleleExpression	1	INTEGER	N/A	SomaticMutation

Class 'GeneCNV'

Instance of this class represents a call on a gene's copy number status as estimated on the basis of a tumor-normal sample pair from a patient. If, for example, we exome sequence 20k genes from a pair of samples, there will be 20k instances.

superclass:	AnalysisResultValue			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisRun	1	INSTANCE	CNVAnalysis	AnalysisResultValue
copyNumber	1	FLOAT	N/A	GeneCNV
copyNumberStatus	1	TEXT	N/A	GeneCNV
gene	1	INSTANCE	Gene	GeneCNV
personStudyId	1	TEXT	N/A	AnalysisResultValue
sampleLongId	1	TEXT	N/A	AnalysisResultValue

Class 'CNVAnalysis'

Instance of this class represents Copy Number Variation analysis performed on the sequencing data produced from one patient's tumor-normal sample pair.

superclass:	TumorVsNormalAnalysis			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin

analysisResultValues	+	INSTANCE	AnalysisResultValue	Analysis
completionTime	1	DATETIME	N/A	Analysis
controlSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis
name	+	TEXT	N/A	Analysis
protocol	1	INSTANCE	AnalysisProtocol	Analysis
runLog	1	TEXT	N/A	Analysis
tumorSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis

Class 'SomaticMutationAnalysis'

Instance of this class represents somatic mutation analysis analysis performed on the sequencing data produced from one patient's tumor-normal sample pair.

superclass:	TumorVsNormalAnalysis			
Attribute name	Cardinality	Value type	Allowed classes	Attribute origin
analysisResultValues	+	INSTANCE	SomaticMutation	Analysis
completionTime	1	TEXT	N/A	Analysis
controlSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis
name	+	TEXT	N/A	Analysis
protocol	1	INSTANCE	AnalysisProtocol	Analysis
runLog	1	TEXT	N/A	Analysis
tumorSample	1	INSTANCE	TissueSample	TumorVsNormalAnalysis