# New concept proposal

# Assay

| | | | |
|---|---|---|---|
| **Author** | Eelke van der Horst,<br>Femke Kopmels | **Date last updated** | 31/10/2023 |
| **Project** | General interest | **Contact** | DCC |
| **Dataset release** | 2024.1 | **Consulted expert** | - |

## 1 Rationale

Assay metadata is essential when sharing (experimental) results: for providing context, ensuring data quality, enabling data integration, and facilitating collaboration and reproducibility in research and clinical settings. An Assay takes a sample and produces data about that sample. For different types of omics research, different types of assays will be relevant, each with their own defining attributes. Therefore, a common *Assay* concept is proposed, that can be used as-is, or inherited by more specific types of *Assay*.

## 2 Comparison to other standards/data models

### 2.1 OBI / EFO

OBI has the 'assay' class, which is "a planned process that has the objective to produce information about a material entity (the evaluant) by examining it" (OBI:0000070). This class is equal to the proposed *Assay* concept. EFO imports OBI's 'assay' class and many of its subclasses (though not all), and defines additional assays. Assay types from both terminologies can be used to type the assay concept (using code). EFO 'Assay' cross references NCIT 'Assay' (NCIT:C60819) defined as "A qualitative or quantitative analysis performed to determine the amount of a particular constituent in a sample or the biological or pharmacological properties of a drug".

### 2.2 EDAM

EDAM includes the procedures used to conduct an experiment as 'Laboratory Techniques' (EDAM: topic_3361). This concept is defined as a 'Topic' rather than an 'Operation', therefore it is considered not relevant for our *Assay* proposal.

## 3 Concept Information

| Concept or concept compositions or inherited | General concept name | General description | Contextualized concept name | Contextualized description | Type | Standard | Value set or subset | Meaning binding | Cardinality for composed Of |
|---|---|---|---|---|---|---|---|---|---|
| concept | Assay | a process with the objective to produce information about a sample by examining it | Assay | a process with the objective to produce information about a  sample by examining it | | | | OBI:0000070 \|assay\| | |
| composedOf | standard operating procedure | standard operating procedure associated to the concept | standard operating procedure | the standard operating procedure that was followed for this assay | Standard Operating Procedure | | | | 0:1 |
| composedOf | predecessor | a preceding process associated to the concept | predecessor | sample processing preceding this assay | Sample Processing | | | RO:0002087 \|immediately preceded by\| | 0:n |
| composedOf | code | coded information specifying the concept | code | code specifying the type of assay | | EFO, OBI or other | for OBI: descendant of: OBI:0000070 \|assay\| | | 1:1 |
| composedOf | identifier | unique identifier identifying the concept | identifier | unique identifier identifying the assay | string | | | | 0:1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| composedOf | start datetime | datetime at which the concept started | start datetime | datetime at which the assay was first executed | temporal | | | | 0:1 |
| composedOf | data file | data file associated to the concept | data file | data file associated to the assay | Data File | | | | 0:n |
| composedOf | sample | any material sample for testing, diagnostic, propagation, treatment or research purposes associated to the concept | sample | material that is being examined by this assay | Sample | | | | 0:n |

| General concept name | Cardinality for concept to Administrative Case | Cardinality for concept to Data Provider | Cardinality for concept to Subject Pseudo Identifier | Cardinality for concept to Source System |
|---|---|---|---|---|
| Assay | 0:n | 1:1 | 0:n | 1:n |

## 4 Impact on the SPHN Dataset

-

## 5 Discussion

As a value set we propose “EFO, descendants of OBI:0000070, or other” to facilitate in foreseen assay types. EFO imports many classes from OBI (such as ‘assay’ and subclasses of assay) but also defines their own. Since EFO is an application ontology for EBI data repositories and OpenTargets, there are also terms missing that are present in OBI but have no purpose for these repositories. For instance, ‘Imaging Mass Cytometry’ (ICM) is available in OBI but not in EFO, while EFO holds a concept for “clone by clone sequencing” that is missing in OBI.

It was discussed to call the proposed concept “(Data) generator”. Although assays such as sequencing are data generators, it can also include processes that do not evaluate a sample, such as a machine or data transformation. Additionally, the *Assay* is more aligned with the existing design of the SPHN model as it provides more specificity with the concept name.

It could be possible that an Assay has more than one Sample as input. This would be common where multiple samples are multiplexed as part of the same sequencing assay. In the case of sequencing assay, the samples can be mapped back via the resulting demultiplexed data file. However, this will not always be the case for all types of assays. Therefore, it was decided to maintain a 0:n cardinality for this more general Assay concept, but subclasses of Assay can be more strict.

When multiple runs are executed for the same Assay, the onset datetime for this concept will be equal to the run datetime of the Run that was first executed.

## 6 Example

Imaging Mass Cytometry (ICM) assay. The output format is .fcs, which is a binary file format with three main segments. This file format is not a descendant of EDAM:format_1915 and therefore assigned the more general format_2333 code for “binary format”.

code: OBI:0003096 |imaging mass cytometry assay|
identifier: example_assay123
start datetime: 2023-06-29
sample:
    identifier: sample_123
    collection datetime: 2023-06-28
    material type code: 127470000 |Tissue specimen from pancreas|
data file:
    format code: EDAM:format_2333 |Binary format