Common Assay Template

# Current

assay title

bioassay type

bioassay

assay format

assay design method

assay supporting method

assay cell line

organism

biological process

target

applies to disease

assay mode of action

result

screening campaign stage

assay footprint

assay kit

physical detection method

detection instrument

perturbagen type

protein identity

gene identity

GO terms

assay sources

related assays

measurement

field

units

operator

threshold

# Future

measurement

multi-measure mode

wavelength, frequencies

signal direction

polarity inversion of signal

reagent data source

commercial site (e.g. Millipore, etc.)

lot numbers

--Biological Material Class: The class of the Biological Material in the assay container (well), relevant to the way the compound accesses the target/focus of the assay. Types would roughly include Physical Chemical Properties (actually no biology, for solubility assessments, etc), Subcellular (lysed cells-for typical Biochemical assays), Whole Cell (for functional assays using cell lines or primary cells), Tissue/Anatomical Structure, Whole Animal.

--Detail associate with some classes above--for instance Whole Cell would indicate cell line or cell type, tissue would indicate tissue type, Whole Animal would indicate species and strain, etc.

--Assay identifiers at source (typically source would be identified, and the internal identifiers at that source used.

--I noticed “multi-measurement” mode is referred to above. Naturally some assays will do multiple measurements from the same well, animal, etc, even when those “reaction vessels” undergo the same experimental conditions and experimental procedures.

--Assay Group/Subassay. There is a general problem with identifying the relationship between tightly related assays, assay groups. An assay (protocol), could be associated with instructions to run an experiment--in other words tightly bound to how experiments are actually run by biologists (sets of wells, plates, animals, etc, in the same “run” on the same days or couple of days). But these assays/protocols may themselves involve wells run with slightly different assay conditions meant to be related together or compared in analysis (i mean aside from the different time points or compound concentrations that are routinely aggregated/reduced to produce IC50s, etc.). Also there can be multiple measurements etc, taken, and multiple stages of the assay (mentioned above.) So something would be useful that connects these multiple parts together, either as connecting multiple assays into a group, or specifying subassays with a single assay. (i suppose these methods are conceptually equivalent).

--Assay “Marker”: the endogenous state or activity that is directly measured by the assay technology. Often that is different from what that state or activity is intended to be a proxy of. For instance, the true assay “focus”/target may be inhibition of GCPR 40, but this may be directly measured by downstream calcium mobilization, which itself might be measured through radioactive tagging, or fluorescence.

--Detection Signal Type (this may be above already): in other words, radioactivity, fluorescence, chemiluminescence, cell counting, visual examination, etc. This is going to be strongly related to raw measurement type, wavelength, frequency, count, etc.

--Normalized measurement type: typically will be things like “Percent Inhibition”, “percent effect”, etc--the result type that comes out after the normalization process--there are differences here too depending on whether measuring results at a single compound concentration, or measuring the results of fitting across concentration series--IC50, Ki, CC50, etc.

--Reference Compounds: used in assay development, or perhaps used in control wells of experiments as proxies for maximum, minimum effects, etc.

--Target Types: room for classification here depending on type of molecular target. Practically all are proteins, but could possibly be DNA, RNA, and even others.

--Protein Target subtype: perhaps most assays involve single proteins, (expressed from single genes), but in many assays (on ion channels for instance), the functional unit that is actually targeted is in fact a complex built from several different proteins--so it would be important to identify the particular complex (many related ion channels can be more precisely identified by the precise protein/genetic identification of the individual components. Similarly, many assays use protein-protein interactions in a similar manner.

--I see now that much of what I have mind-dumped above is already in the lower levels of the categories already included. However, in general I’m not sure that the structure of these hierarchies picks out what is most important in the assays in the best way. I think there could be a more strongly indicated distinction between parameters/categories that are most fundamental, and most important to the scientific analysis of the protocols, and those elements that are tied to particular methods and technologies, but are not of primary importance scientifically. After all, many of the methods and technologies above, theoetically, are supposed to be “irrelevant” to the final results--normalized out of existence. (of course we know in practice that is not true--nevertheless the details of technology should still be of secondary importance).

--are splice variants indicated in the genetic identifications associated with the target? Often the assays are meant to see results only for specified splice variants. Or also meant to target proteins in particular states (eg. phosphorylation states), or in particular conformations (eg. ion channel conformations)

--is species pulled out to identify both biological material (eg. cell line), and target species. Naturally with transfections these can often be different.

--also, even if a cell line is not used directly in an experiment, a particular cell line (often an insect cell line), is used as an expression system to produce a protein that might be used in solution in an experiment. Should the expression system be identified separately?

--the various result sets coming out of an assay need to be identified according to various orthogonal dimensions: are the sets raw or normalized, which measures are they associated with, what stage of an assay are the associated with (screening, confirmation), what is the level of the data reduction process (single points at specific compound concentrations, vs fitted data across a series), etc.

--what settings/parameters were used for data fitting? Were there parameters, thresholds used to assess confirmations? What parameters were used to assess data integrity, outliers? Should these settings be recorded?

publication source

Nucleic Acid information -- RNA/DNA, oligos: MW, structure, sequence, modifications?

... HELM code

Replicated/”close” entry handling - (HepG cells, vs Hep-G cell)

Expanded cell line information - patient-type source, vendor source, mutation(s), passage number if relevant

Animal model lines too

link-backs

confirmatory assay of

(etc.)

# New Template Types

# FAIRness Improvements

Initial evaluation shown below...

###### **Questions answers of your last evaluation**

|  |  |
| --- | --- |
| Q0: What is the URL of the resource to be evaluated? | http://beta.bioassayexpress.com |
| Q1: Do you have a URL to a registered scheme that defines the globally-unique structure of the identifier(s) for your digital resource? | V |
| Q2: Do you have a URL to a document that defines the persistence policy of your identifier(s)? | X |
| Q3: Do you have a URL to a document that contains machine-readable metadata for the digital resource? | V |
| Q4: Do you have a URL for the file format of this metadata? | X |
| Q5: Do you have a URL to the metadata document that contains the globally unique and persistent identifier for the digital resource? | V |
| Q6: Do you have a URL to the data described by in that metadata document? | V |
| Q7: Is there a URL to a search engine that can find your data? | V |
| Q8: Is there a query that when executed will discover your RESOURCE ID (found in the first page of the search)? | V |
| Q9: Do you have a URL to the description of the access protocol? | X |
| Q10: The access protocol is Open (Yes/No)? | V |
| Q12: Authorization is required to access the content of my RESOURCE ID (Yes/No)? | X |
| Q13: If Q12 was "Yes”, do you have a URL that resolves to a description of the process to obtain access to restricted content? | ? |
| Q14: Do you have a URL to a metadata longevity plan? | X |
| Q15: Do you have a URL to the specification of a knowledge representation language? | X |
| Q16: Do you have one or more (max 3) URLs representing the vocabularies used within the (meta)data that is returned by resolving the RESOURCE ID? | V |
| Q17: Do you have a URL to a formal Linkset (defined at: https://www.w3.org/TR/void/#linkset ) that describes at least a portion of the content at RESOURCE ID? | X |
| Q11: The access protocol (royalty) free (Yes/No)? | V |
| Q18: Do you have a URL for your usage license document for the content returned from RESOURCE ID (ideally for BOTH (independently) the data and its associated metadata)? | V |
| Q19: Do you have URLs (maximum 3) to the vocabularies being used to describe the provenance of the content resolved from RESOURCE ID (be that data, or metadata; e.g. http://purl.org/dc/terms/ )? | V |
| Q20: Do you have a URL (maximum 3) for the vocabularies being used to describe the domain information of the content resolved from RESOURCE ID (be that data, or metadata)? | V |
| Q21: Do you have a URL that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID is compliant with the standards of your community? | X |