

Assay annotation with ontologies

Introducing the DataFAIRy project

Alex M. Clark

alex@collaborativedrug.com



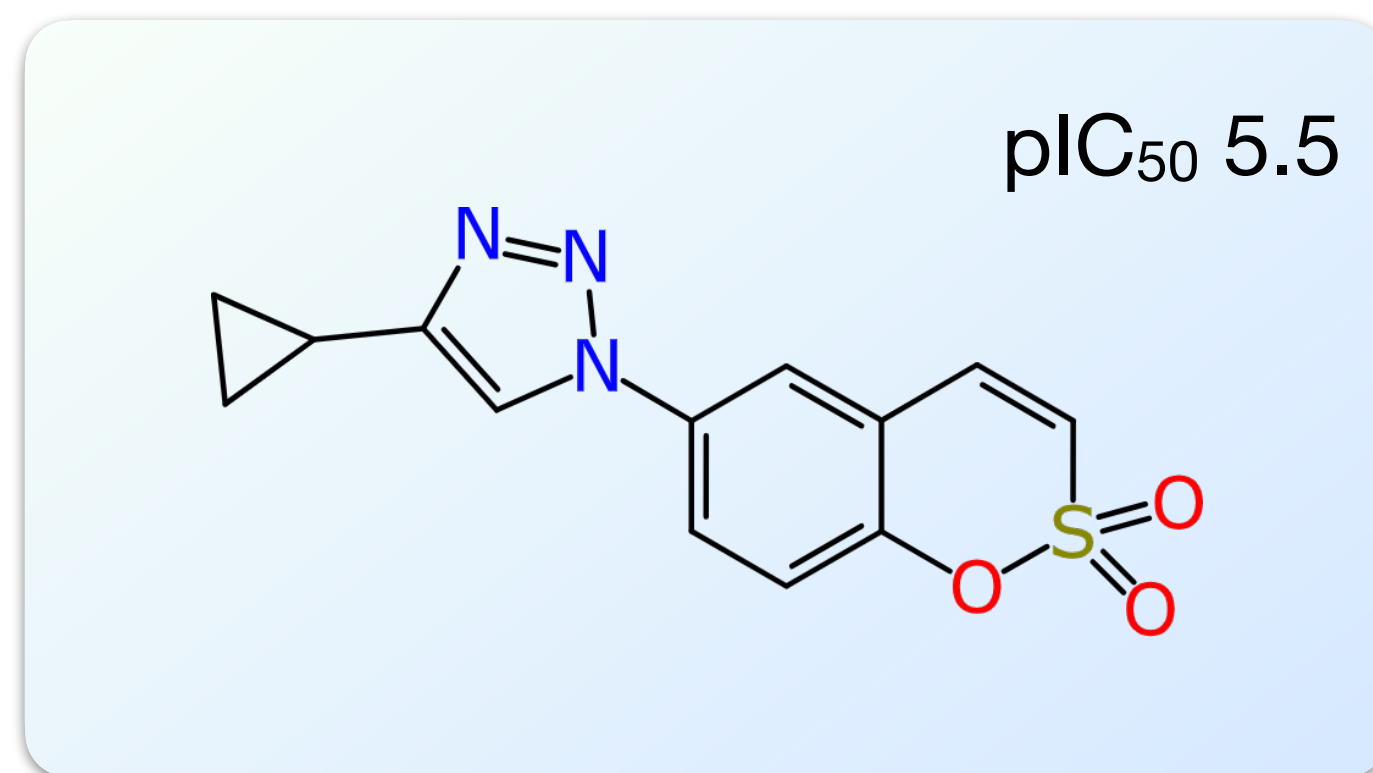
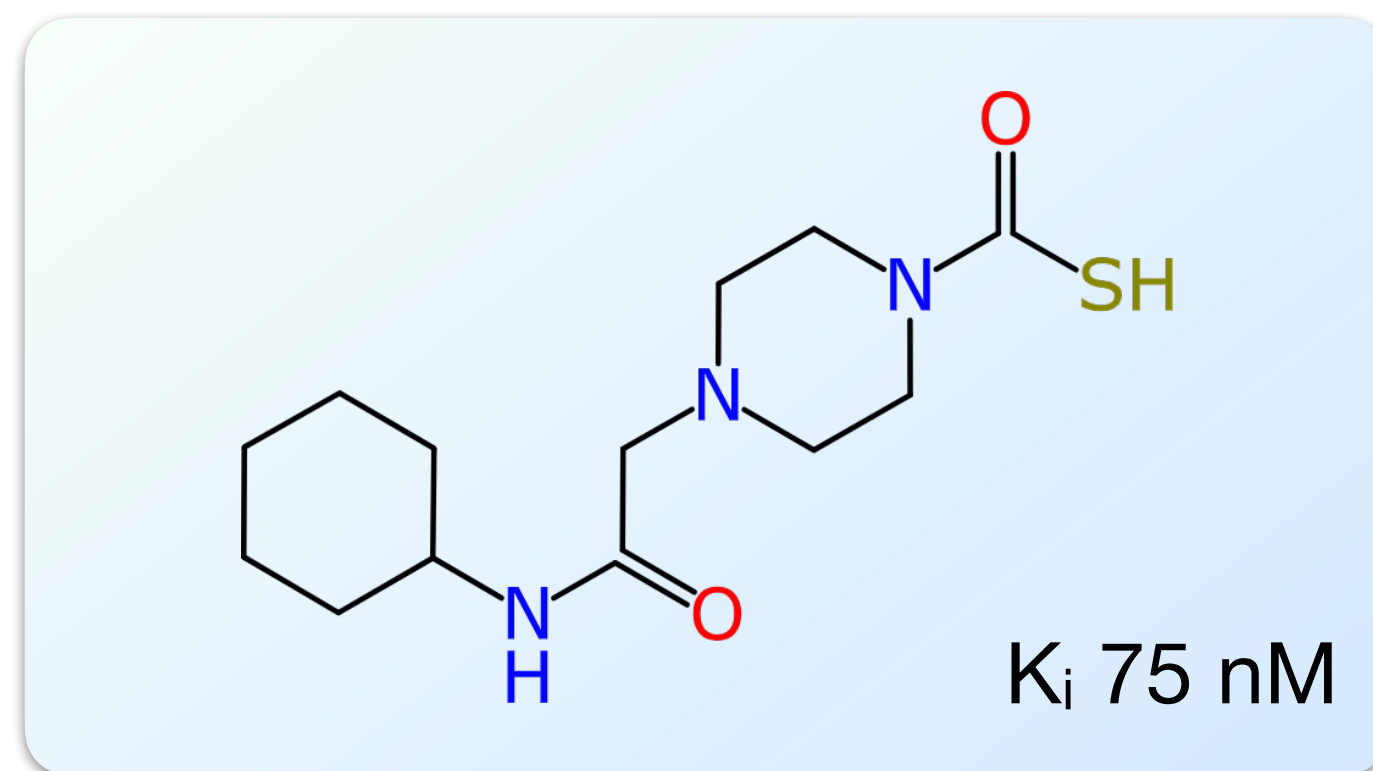
CDD.VAULT[®]
Complexity Simplified



Background

- ❖ cheminformatics and bioinformatics: machine readable for decades
- ❖ ... assay informatics under-developed: measurements are rarely FAIR

machine
readable



Cell-Free Homogeneous Primary HTS to Identify Inhibitors of GSK3 β Activity

- (1) Dispense 1 μ L/well of CABPE, 0.5 μ L of ATP, and 1 μ L of positive control (WB510 or AB) in respective wells according to plate design to 1536-well assay ready plates (Aurora 29847) that contain 2.5 nL/well of 10 mM compound using BioRAPTR (Beckman) to start the reaction. Incubate at room temperature for 60 minutes.
- (2) Add 2.5 μ L/well of ADP-glo (Promega, V9103) with BioRAPTR, incubate at room temperature for 40 minutes
- (3) Add 5 μ L/well of ADP-glo (Promega, V9103) with Combi nL (Thermo), incubate at room temperature for 30 minutes

The GSK3 β Assay Kit is designed to measure GSK3 β activity for screening and profiling applications using Kinase-Glo[®] (Promega) as a detection reagent. The GSK3 β Assay Kit comes in a convenient 96-well format, with enough purified recombinant GSK3 β enzyme, GSK3 β substrate (GSK substrate peptide), ATP, and kinase assay buffer for 100 enzyme reactions.

Assay Kit Format: Luminescent

Materials Required
Kinase-Glo[®] Max Assay (Promega #V6071)
Dithiothreitol (DTT, 1 M; optional)
Microplate reader capable of reading luminescence
Adjustable micropipettor and sterile tips
30°C incubator

not so
much...



Ontologies to the rescue

- ❖ Public, standardized, well defined, but... how to use them?





Templates

❖ **Common Assay Template:** most of what you need to know for screening

bioassay type

target

detection instrument

bioassay

applies to disease

perturbagen type

format

mode of action

protein identity

design method

result

gene identity

supporting method

campaign stage

GO terms

cell line

footprint

sources

organism

assay kit

biological process

physical detection method

BioAssay templates for
the semantic web

[10.7717/peerj-cs.61](https://doi.org/10.7717/peerj-cs.61)

❖ **Pistoia Alliance Assay Template:**
larger and more detailed

Bioassay Protocol Metadata Annotation:
Proposed Standards Adoption

osf.io/preprints/osf/pz8u7

BioAssay Express



- ❖ Built for original curation of public data
- ❖ Test-bed for many experimental features
 - ◆ data entry
 - ◆ searching
 - ◆ analysis
 - ◆ machine learning
 - ◆ text creation
 - ◆ (among others)

The screenshot displays the BioAssay Express web interface for editing an assay (ID: 7008). The interface is divided into several sections:

- Protocol Text:** Contains the assay title "HepG2 Cytotoxicity Assay Measured in Cell-Based System Using Plate Reader - 7071-02_Inhibitor_Dose_DryPowder_Activity_Set10" and keywords: "Cytotoxicity Assay, luminescence, CellTiter-Glo, HepG2".
- Assay Annotations:** A list of key-value pairs for various assay parameters, such as "assay title", "bioassay type" (functional), "assay format" (cell based format), "assay design method" (ATP quantitation using luciferase), "assay cell line" (Hep-G2 cell), "organism" (Homo sapiens), "biological process" (cell population proliferation), "assay mode of action" (growth inhibition), "result" (AC50), "screening campaign stage" (compound toxicity assay), "assay footprint" (384 well plate), "assay kit" (CellTiter-Glo Luminescent Cell Viability Assay), "physical detection method" (bioluminescence), "detection instrument" (EnVision Multilabel Reader), "perturbagen type" (unknown), "protein identity", "gene identity", "GO terms", "assay sources" (Broad Institute (Harvard-MIT)), "related assays" (AID 623896), "measurement" (AC50_uM), and "units" (micromolar).
- Similar Assays:** A list of related assay IDs (AID) such as AID 651898, AID 652117, AID 652118, AID 720585, AID 720594, AID 743185, AID 743354, AID 743358, AID 624285, and AID 651860.



Data Curation

- ❖ At **Collaborative Drug Discovery**:
 - ◆ grant funded: expert curation + machine learning support
 - ◆ 4000 assays from **PubChem** using **Common Assay Template**
- ❖ **Pistoia Alliance DataFAIRy** project:
 - ◆ 2300 detailed assays using **Pistoia Alliance Assay Template**
 - ◆ diverse sources: literature, PubChem, vendor collections
- ❖ Openly available:
 - ◆ <https://beta.bioassayexpress.com>
 - ◆ <https://github.com/cdd/bioassay-express>



What can you do with marked up assays?

❖ Search by narrowing with *exact* and *precise* criteria

The screenshot displays a search interface for bioassays with the following components:

- LAYER 1: bioassay type**
 - Keywords:
 - empty ?
 - with text ?
- LAYER 2: organism**
 - Keywords:
 - empty ?
 - with text ?
 - organism ?
 - cellular organisms ?
- Assay Format Filter**
 - assay format
 - biochemical format
 - protein format
 - protein complex format
 - single protein format
 - cell based format
 - cell-free format
 - plasma format
 - subcellular format
 - cell membrane format
 - cytosol format
 - whole cell lysate format
 - organism-based format
 - tissue-based format
- Organism Filter**
 - protozoan ?
 - Plasmodium falciparum ? (1 of 47)
 - Plasmodium falciparum 3D7 ? (2 of 3)
 - prokaryote ?
 - virus ?
 - absence ?
- Assay Cell Line Filter**
 - HEK-293 cell
 - HEK-293T cell
 - cell line cell
 - immortal cell line cell
 - CHO-K1 cell
 - immortal dento-alveolar joint-derived cell line cell
 - immortal mouse dento-alveolar joint-derived cell line cell
 - Neuro-2a cell
 - immortal female gonad-derived cell line cell
 - CHO cell
 - absence
 - not applicable
 - organism
 - cellular organisms
 - eukaryote
 - metazoa
 - vertebrate
 - mammalian
 - Bos taurus
 - Homo sapiens
 - Mus musculus
 - Rattus norvegicus
- Assay Biology Component Filter**
 - biological process
 - biological regulation
 - regulation of biological process
 - regulation of cellular process
 - regulation of cell communication
 - regulation of signal transduction
 - regulation of growth
 - regulation of cell growth

❖ Can also use annotations like fingerprints



Annotation as a product

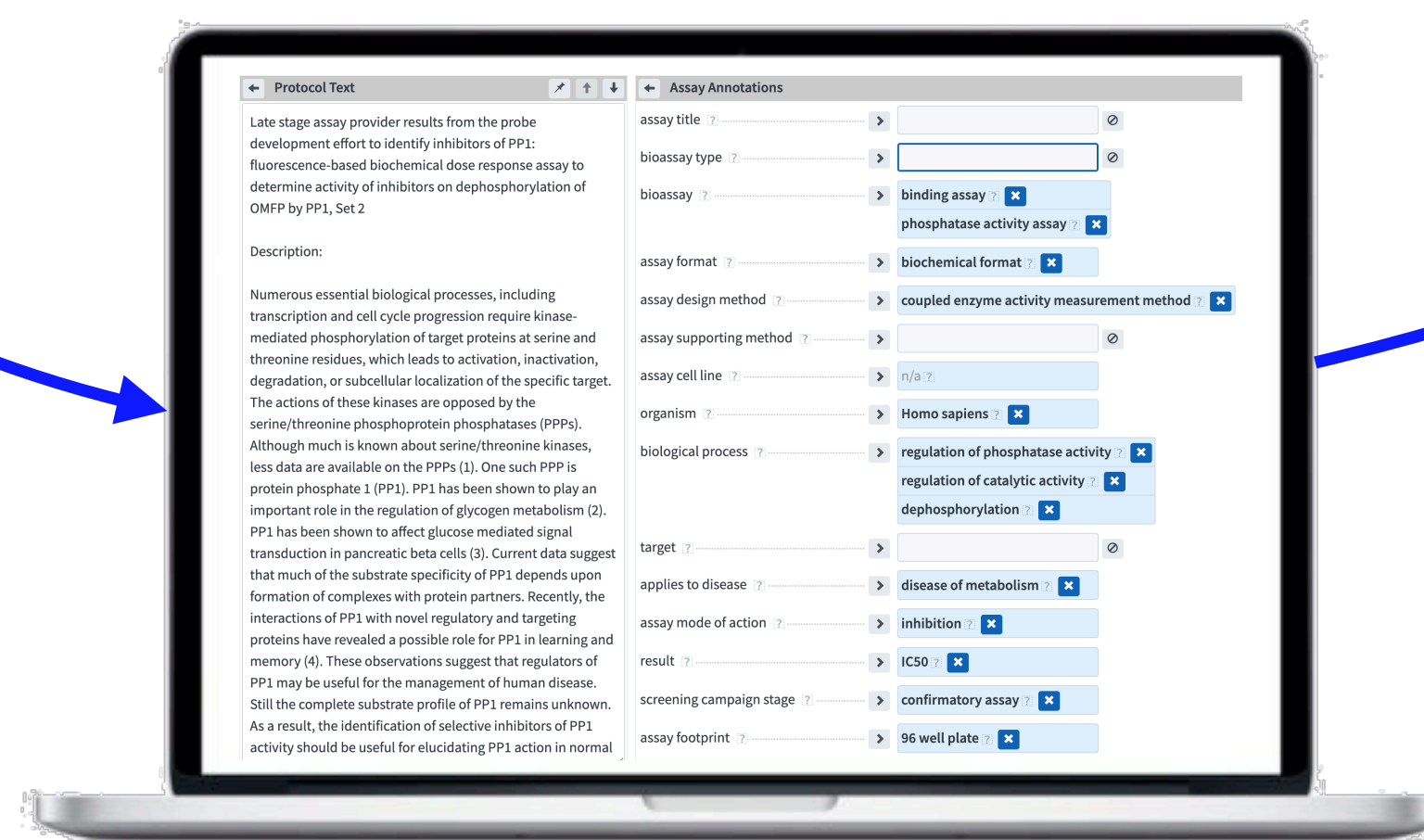
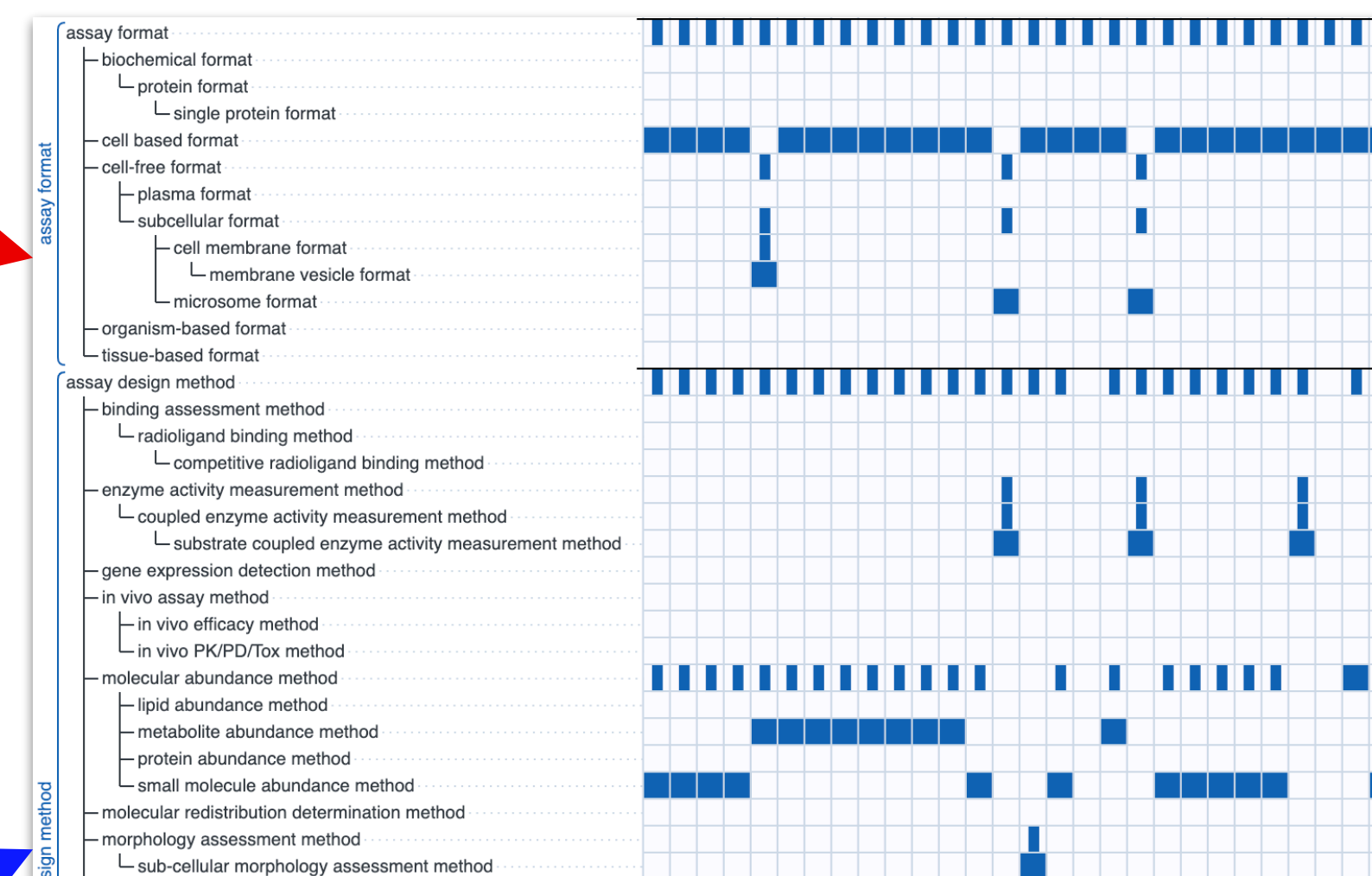
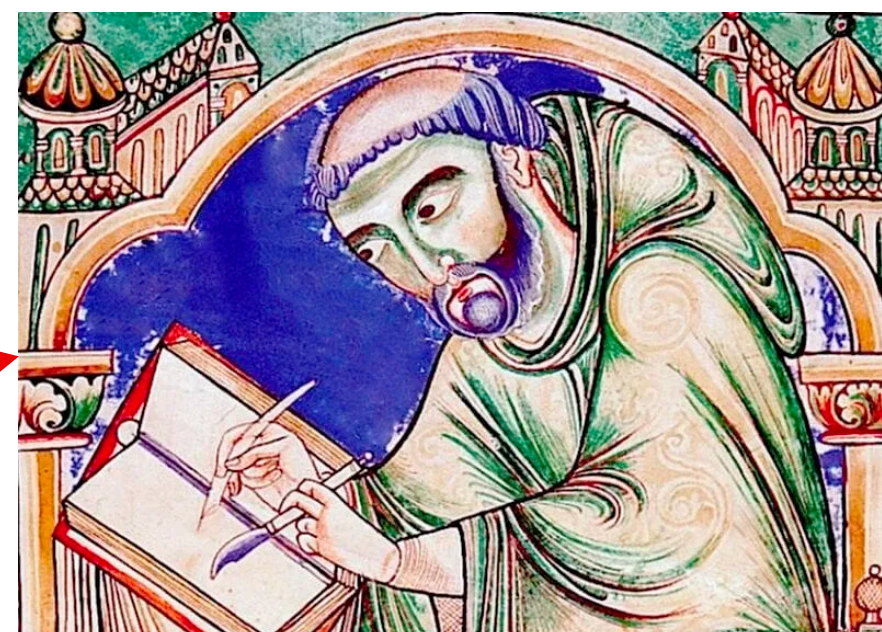
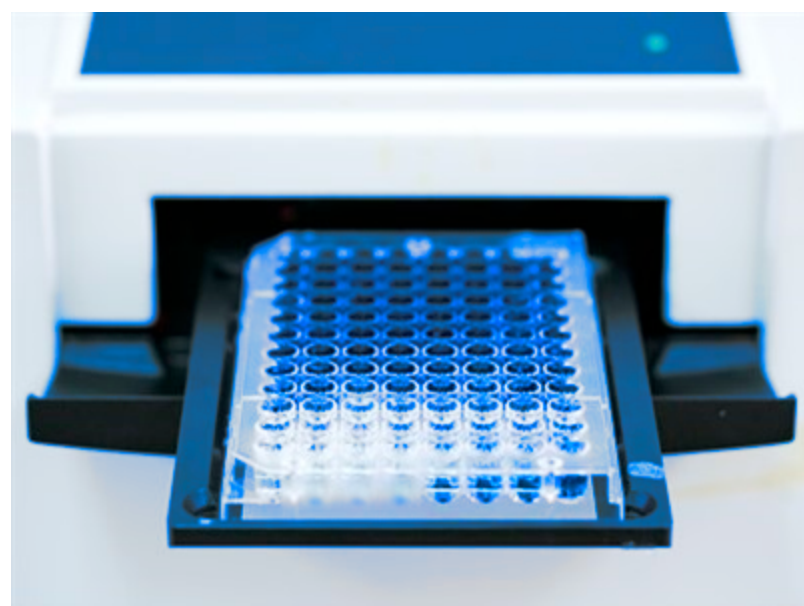
- ❖ **BioAssay Express**: advanced prototype, open source
- ❖ **CDD Vault Annotations**: 80/20 rule, same data model
- ❖ Vault is used daily by thousands of scientists, meet them where they're at
- ❖ Knowing about the *semantic web* is optional
- ❖ Introduce ontologies and templates gently
 - ◆ well defined set of terms
 - ◆ everyone has access to same vocabulary
 - ◆ like what they had before, except better

The screenshot displays the CDD Vault Probes Vault interface. The top navigation bar includes 'CDD.VAULT · Probes Vault', 'Help', and 'Log out'. Below this is a secondary navigation bar with 'Explore Data', 'Import Data', 'Reports', 'Settings', and a user profile 'Alex Clark'. The main content area features a search bar with the text 'Search molecules...' and a 'Go' button. A sidebar on the left lists various data sources with counts, such as 'Projects' (0), 'ADMEdata.com for Ekins Ca...' (0), 'Alex Clark Sandbox' (0), 'AZ Public ChEMBL Data' (0), 'Biocom PII Comparisons' (0), 'CDD Advocate Group' (0), 'MIT Sandbox Vault' (0), 'NCATS ASPIRE' (0), 'Registration System Demo V...' (0), 'Samantha Jeschonek Demo ...' (0), and 'Public Data' (1). The main area shows a list of molecules, with the first one highlighted: 'JHU- 12792' (CDD-2655024) with 'Synonyms: (no synonyms)' and 'Protocols: 5'. Other molecules listed include JHU- 12791 (CDD-2654766), JHU- 12790 (CDD-2626575), JHU- 12789 (CDD-2654327), JHU- 12788 (CDD-2655031), and JHU- 12787 (CDD-2654491). Each molecule entry includes a chemical structure and associated metadata. The bottom of the page shows the URL 'https://app.collaborativedrug.com/vaults/4126/molecules/4758311'.



Curation: now vs. then

❖ Writing up text, then re-deriving machine readable description... ?



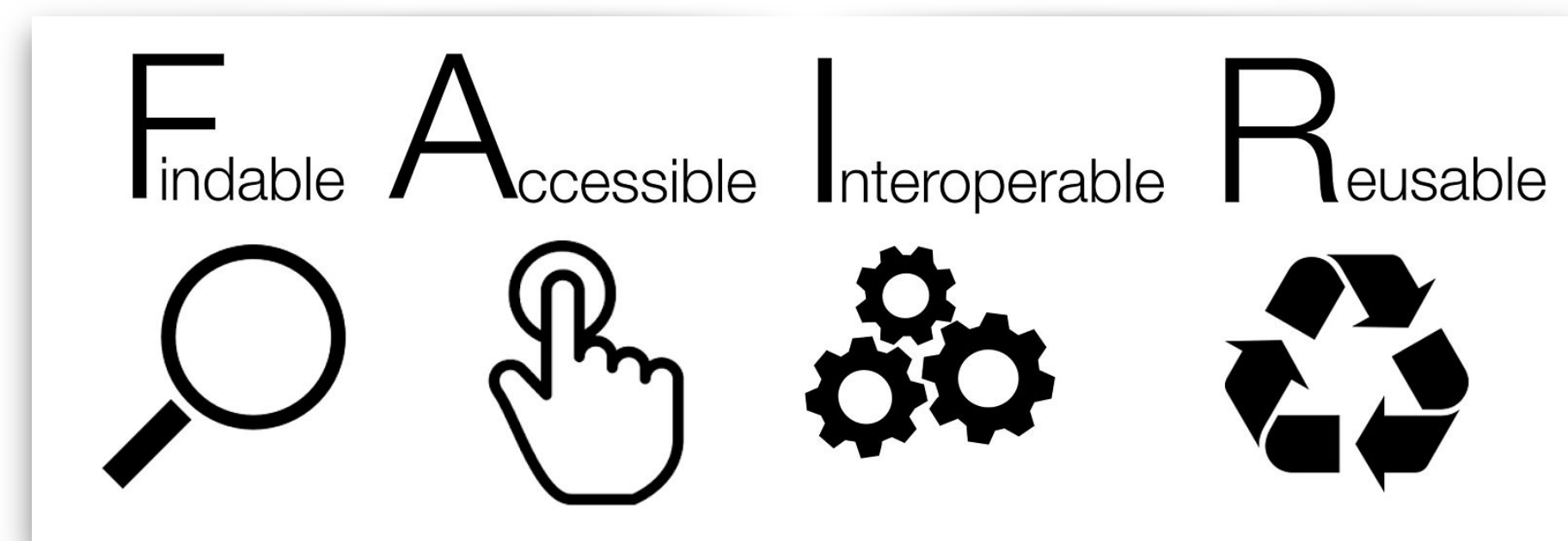
❖ Make the FAIR UI *easier*

❖ Introduce short term payoffs

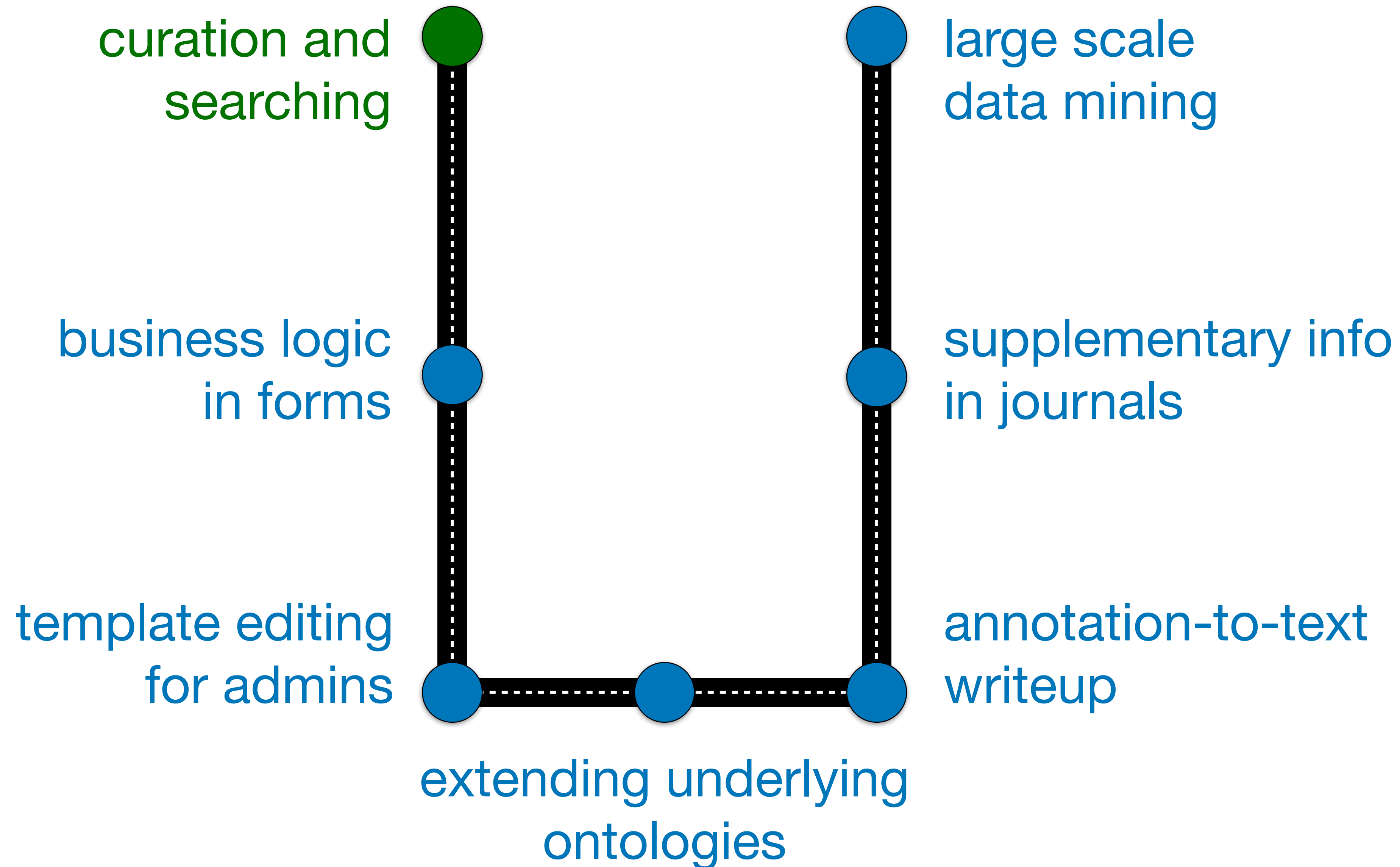


Collaboration

- ❖ Common terms: ontologies are universal
- ❖ Assays are directly comparable, regardless of source
- ❖ Interoperability:
 - ◆ reduce/eliminate ambiguity
 - ◆ share assays with CROs (and acquisitions)
 - ◆ preserve institutional knowledge
- ❖ Completeness: you know what you know...



Roadmap



Questions?



- ❖ Contact:

- ◆ Alex M. Clark alex@collaborativedrug.com (Collaborative Drug Discovery)

- ❖ Thanks to the Vault & BioAssay Express teams