

Using ontologies to make bioassay protocols machine readable

Alex M. Clark

alex@collaborativedrug.com



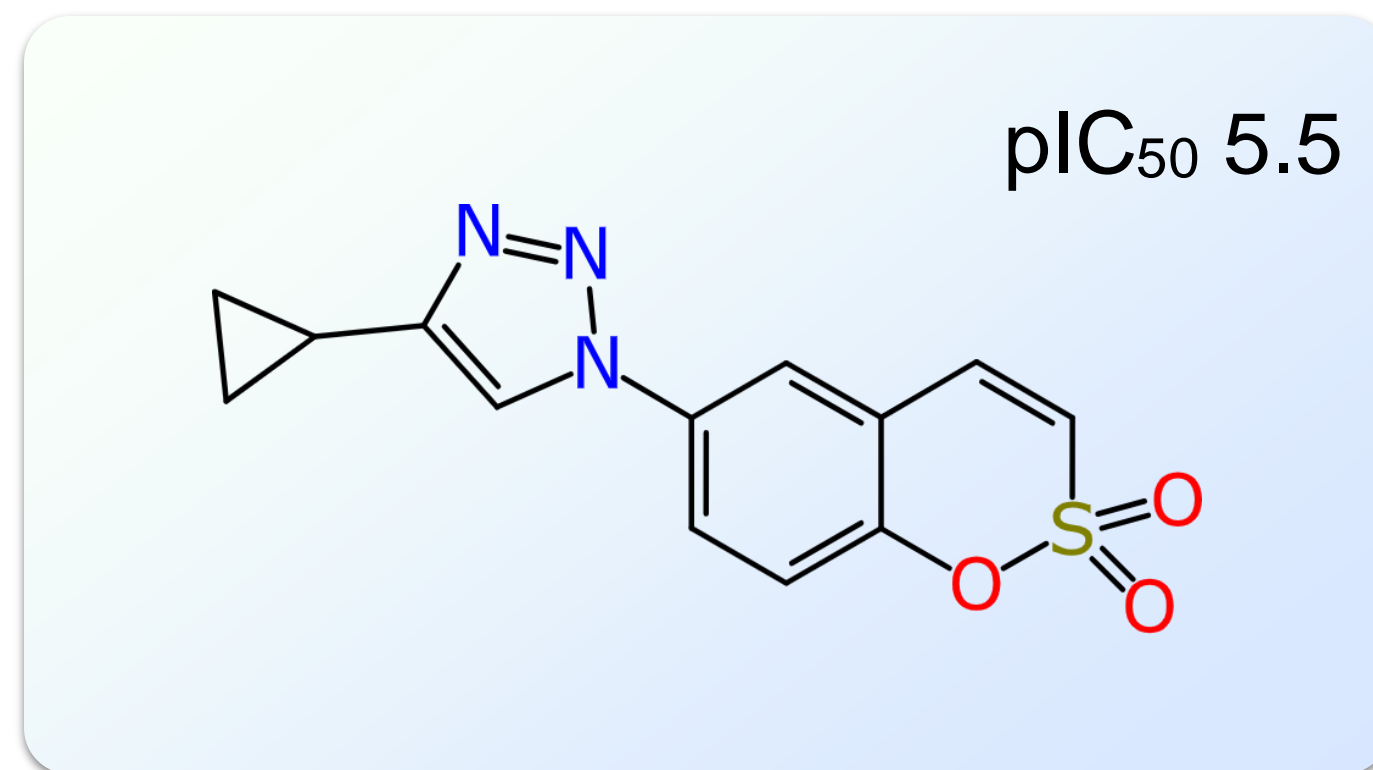
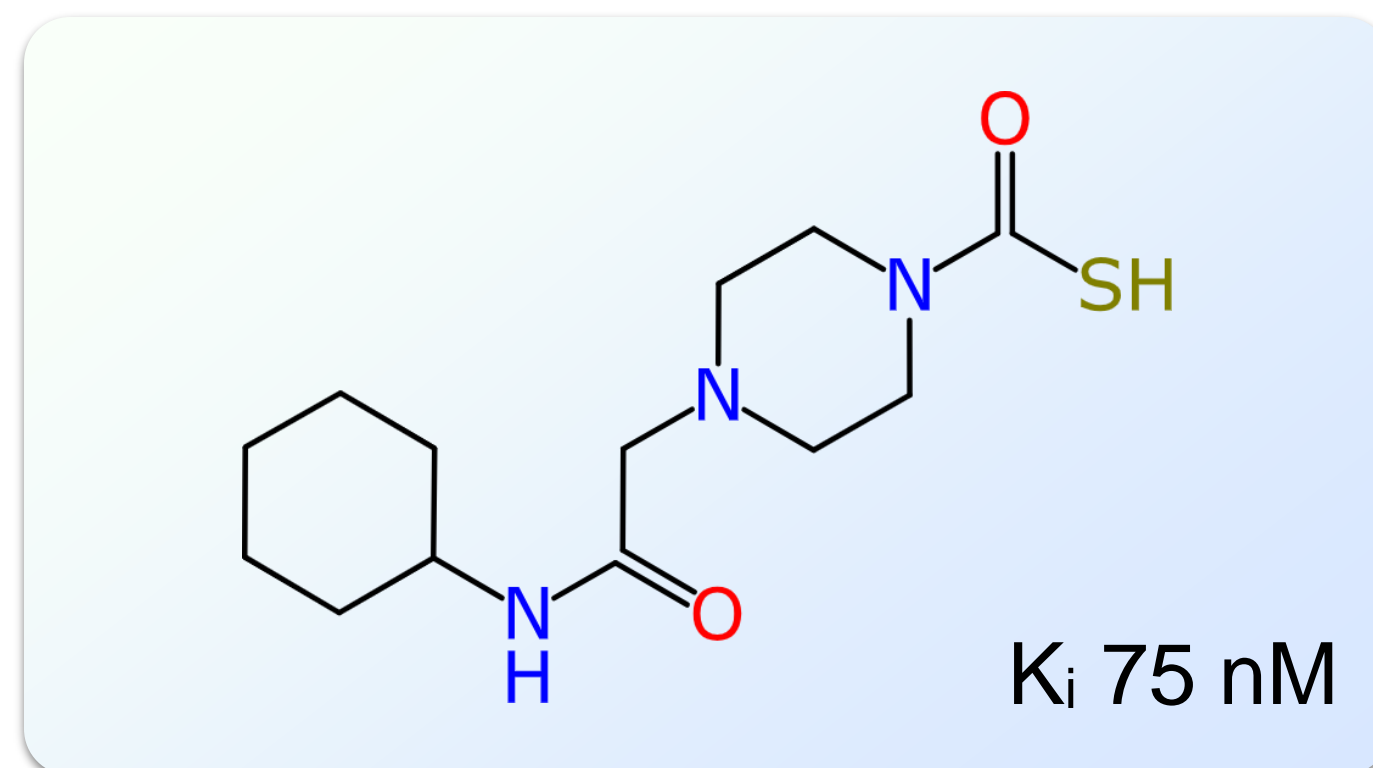
CDD.VAULT[®]
Complexity Simplified



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 GSK3β Activity

- (1) Dispense 1 μL/well of CABPE, 0.5 μL of ATP, and 1 μL of positive control GW8510 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 μL/well of ADP-glo (Promega, V9103) with BioRAPTR, incubate at room temperature for 40 minutes
- (3) Add 5 μL/well of ADP-glo (Promega, V9103) with Combi nL (Thermo), incubate at room temperature for 30 minutes

The GSK3β Assay Kit is designed to measure GSK3β activity for screening and profiling applications using Kinase-Glo® (Promega) as a detection reagent. The GSK3β Assay Kit comes in a convenient 96-well format, with enough purified recombinant GSK3β enzyme, GSK3β 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...

Current best practices

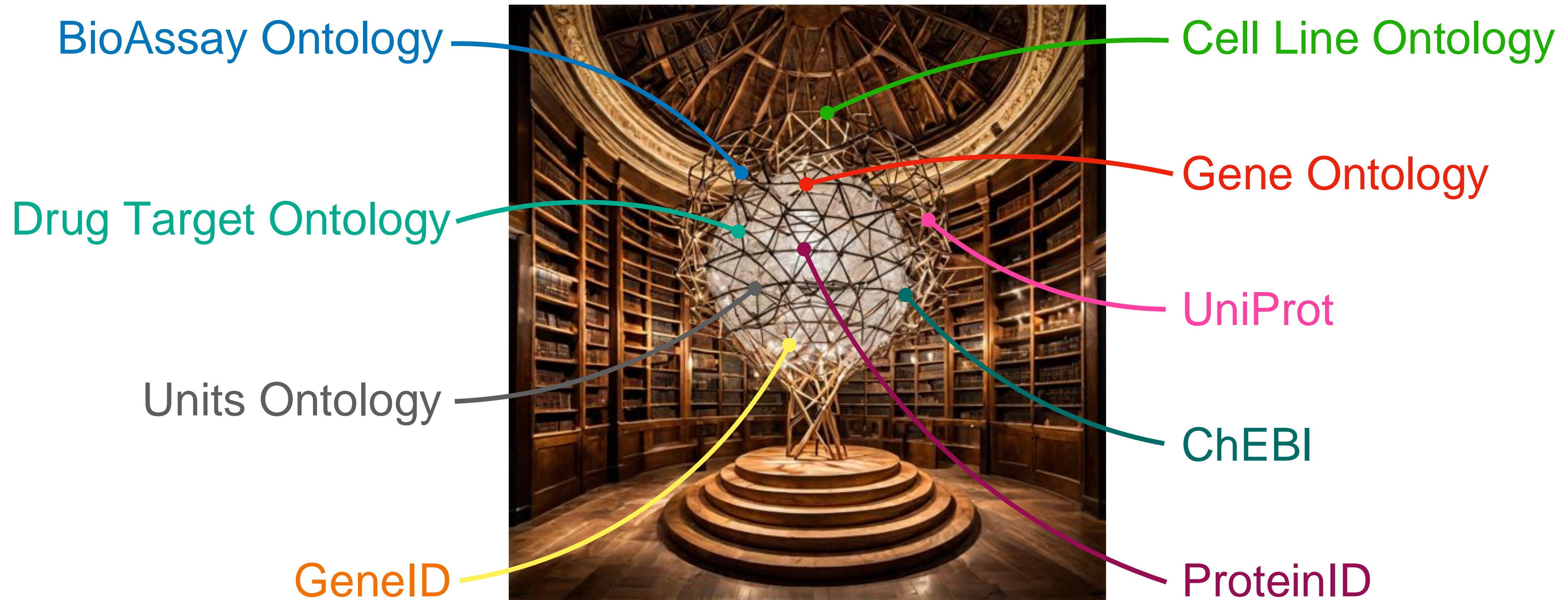


- ❖ Institutions often have partially systematic assay protocol ELNs
 - ✦ e.g. *organism* or *protein target* has well defined dictionary
 - ✦ other data mostly made up of
 - **free text**
 - **pick lists**
 - **mnemonic titles**
 - **Word docs**
- ❖ Consistency between institutions is rare
- ❖ Publications usually terse, inter-referenced or implied

Ontologies to the rescue



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





Ontology templates

- ❖ Intermediate step:
 - ◆ select ontologies relevant to the domain
 - ◆ cherry-pick branches and place them into categories
- ❖ **Common Assay Template:** most of what you need to know for screening

bioassay type

bioassay

format

design method

supporting method

cell line

organism

biological process

target

applies to disease

mode of action

result

campaign stage

footprint

assay kit

physical detection method

detection instrument

perturbagen type

protein identity

gene identity

GO terms

sources

BioAssay templates for
the semantic web
[10.7717/peerj-cs.61](https://doi.org/10.7717/peerj-cs.61)

Template editor



Edit Ontology Template

Schema Prefix:

common assay template

- + bioassay type
 - bioassay
 - bioassay
 - assay kit
 - experimental specification
 - bioassay specification
 - measure group
 - measure group specification
- + 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

Undo Redo

Edit Value

URI

- + absence
- assay bioassay
- + assay kit
- bioassay
 - + cell cycle
 - + cell growth
 - + cell morphology
 - cell motility
 - cell permeability
 - chaperone activity
 - + epigenetic assay
 - genotoxicity assay
 - localization assay
 - + membrane potential assay
 - + metabolomic assay
 - metastasis assay
 - organism behavior assay
 - oxidative phosphorylation assay
 - oxidative stress assay
 - + pharmacodynamic assay
 - + pharmacokinetic assay
 - + physicochemical assay
 - protein folding assay
 - protein profiling assay
 - protein stability assay
 - protein unfolding assay
 - protein-turnover assay

Name

Description

Item Whole Branch Container Exclude Exclude Branch Parent URI

CANCEL

Pick Term: bioassay

bioassay

- + cell cycle assay
- + cell growth assay
- + cell morphology assay
- cell motility assay
- cell permeability assay
- chaperone activity assay
- + epigenetic assay
- genotoxicity assay
- localization assay
- + membrane potential assay
- + metabolomic assay
- metastasis assay
- organism behavior assay
- oxidative phosphorylation assay
- oxidative stress assay
- + pharmacodynamic assay
- + pharmacokinetic assay
- + physicochemical assay
- protein folding assay
- protein profiling assay
- protein stability assay
- protein unfolding assay
- protein-turnover assay
- redistribution assay
 - dye redistribution assay
- ion redistribution assay
 - calcium redistribution assay
 - metal ion redistribution assay
 - protein redistribution assay
- + second messenger redistribution assay
- rna splicing assay
- + safety pharmacology assay
- sensitizer assay
- toxicity assay
- + transporter assay
- viral titer assay

Calcium functions as second messenger in cell signaling.

Search

CLOSE



Body of practice

- ❖ Use the template to create real data: set a strong precedent
- ❖ Grant funded project (*ca.* 2014):
 - ◆ create hybrid curation system: NLP & ML + expert curation
 - ◆ accelerate (rather than replace) scientists
 - ◆ gets easier and quicker as more assays are annotated
- ❖ Curated ~4000 assays from PubChem MLPCN (Molecular Libraries)
- ❖ Data can be found at:
 - ◆ <http://beta.bioassayexpress.com>
 - ◆ <https://github.com/cdd/bioassay-express>

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 shows the BioAssay Express web interface. The main content area is divided into three sections: Protocol Text, Assay Annotations, and Similar Assays. The Protocol Text section contains the following text:

HepG2 Cytotoxicity Assay Measured in Cell-Based System Using Plate Reader - 7071-02_Inhibitor_Dose_DryPowder_Activity_Set10

Keywords: Cytotoxicity Assay, luminescence, CellTiter-Glo, HepG2

Assay Overview: The cytotoxicity assay uses wild-type HepG2 cells from ATCC test for cytotoxicity. The assay measures cellular ATP levels as a surrogate marker of cell viability with Pekin-Elmer CellTiter-Glo. Compound are tested at a range of concentrations.

Expected Outcome: Compounds identified as actives were toxic to cells at a compound concentration less than 10 uM. Activity in the assay leads to a reduction in cellular ATP levels which correlates with a decreased luminescence signal from the read reagent (CellTiter-Glo) and indicates cytotoxicity. Compounds that exhibit no cytotoxicity at >30 uM will be prioritized for additional studies.

PRESENCE OF CONTROLS: Neutral control wells (NC; n=32) and positive control wells (PC; n=32) were included on every plate.

EXPECTED OUTCOME: Active compounds result in decreasing readout signal.

Autogenerated Text

This is a compound toxicity assay investigating the biological process of cell population proliferation, in Homo sapiens.

This is a cell viability ATP quantitation assay, functional in a cell based format, using an ATP quantitation using luciferase, with the assay kit CellTiter-Glo Luminescent Cell Viability Assay. The cell line Hep-G2 cell was used. It was conducted in 384 well plates, with the detection method of bioluminescence, using an EnVision Multilabel Reader. This assay tested the mode of action of growth inhibition by unknown perturbagens.

The Assay Annotations panel lists the following fields and values:

- assay title: "HepG2 Cytotoxicity Assay Measured in Cell-Based System Using Plate Reader - 7071-02_Inhibitor_Dose_DryPowder_Activity_Set10"
- bioassay type: functional
- bioassay: cell viability ATP quantitation assay
- assay format: cell based format
- assay design method: ATP quantitation using luciferase
- assay supporting method: (empty)
- assay cell line: Hep-G2 cell
- organism: Homo sapiens
- biological process: cell population proliferation
- target: n/a
- applies to disease: n/a
- 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: (empty)
- gene identity: (empty)
- GO terms: (empty)
- assay sources: Broad Institute (Harvard-MIT)
- related assays: AID 623896
- measurement: (empty)
- field: AC50_uM
- units: micromolar

The Similar Assays sidebar on the right lists the following IDs:

- AID 651898
- AID 652117
- AID 652118
- AID 720585
- AID 720594
- AID 743185
- AID 743354
- AID 743358
- AID 624285
- AID 651860

What can you do with marked up assays?



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

The screenshot displays a search interface for bioassays. It features several filter layers:

- LAYER 1: bioassay type**
 - Keywords:
 - empty ?
 - with text ?
- LAYER 2: organism**
 - Keywords:
 - empty ?
 - with text ?
 - organism ?
 - cellular organisms ?

The main visualization is a heatmap where rows represent different assay categories and columns represent individual assays (AID numbers). The categories include:

- 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
- bioassay type
 - ADMET
 - binding type
 - functional
 - bioassay
 - binding assay
 - protein-DNA interaction assay
 - protein-protein interaction assay
 - protein-small molecule interaction assay
 - radioligand binding assay
 - enzyme activity assay

Below the heatmap, there are additional filters for organism and assay biology component:

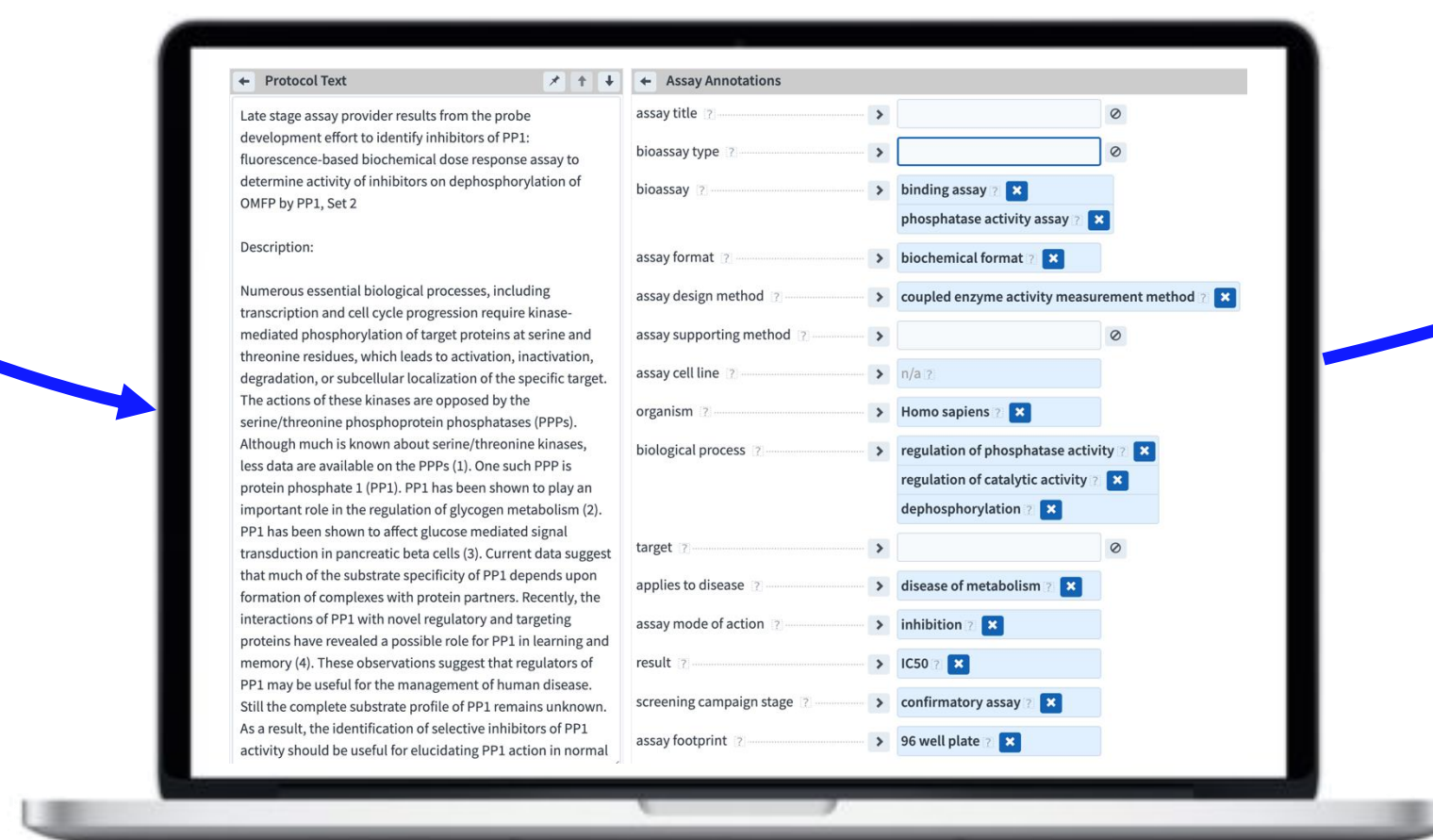
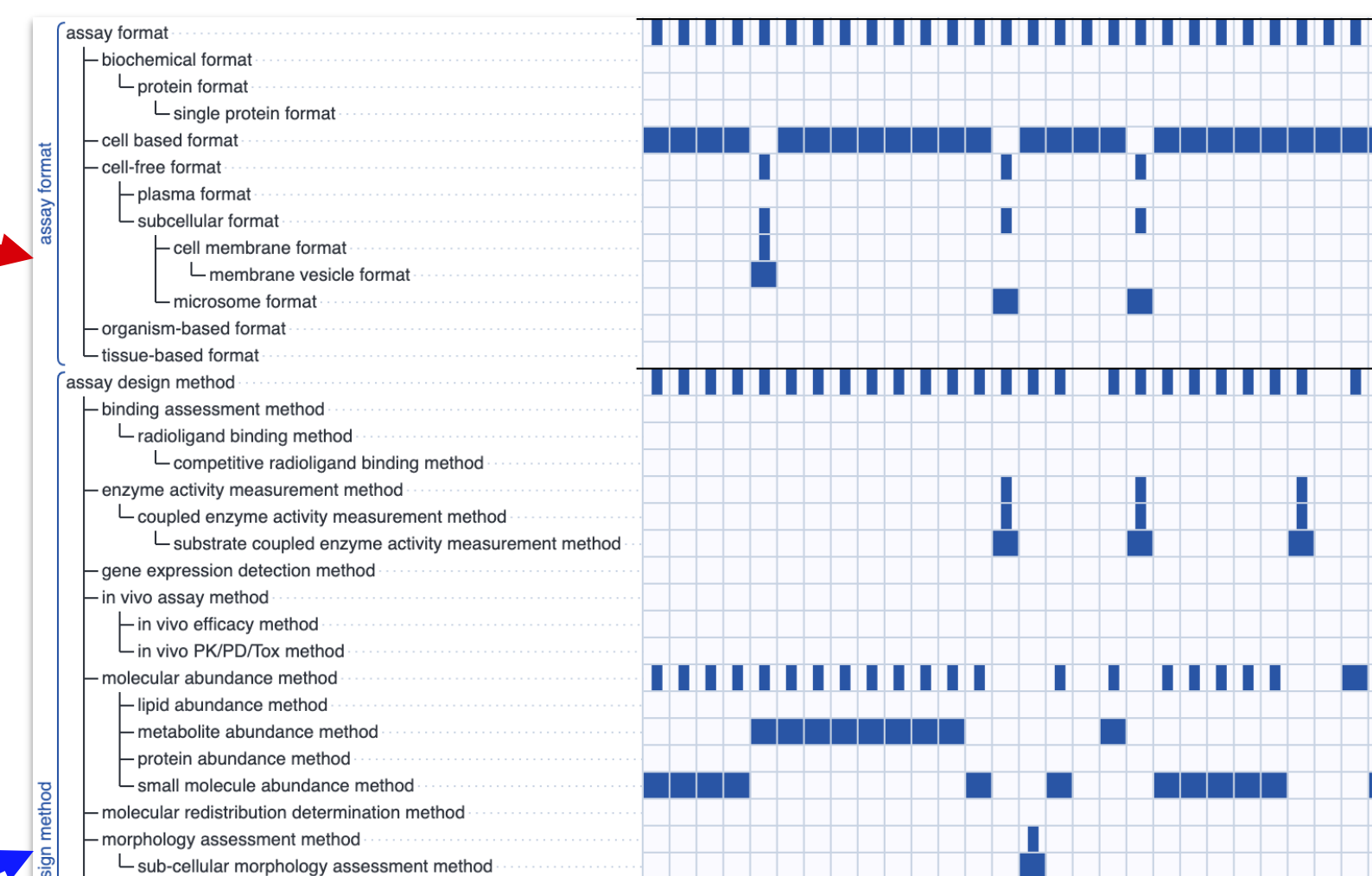
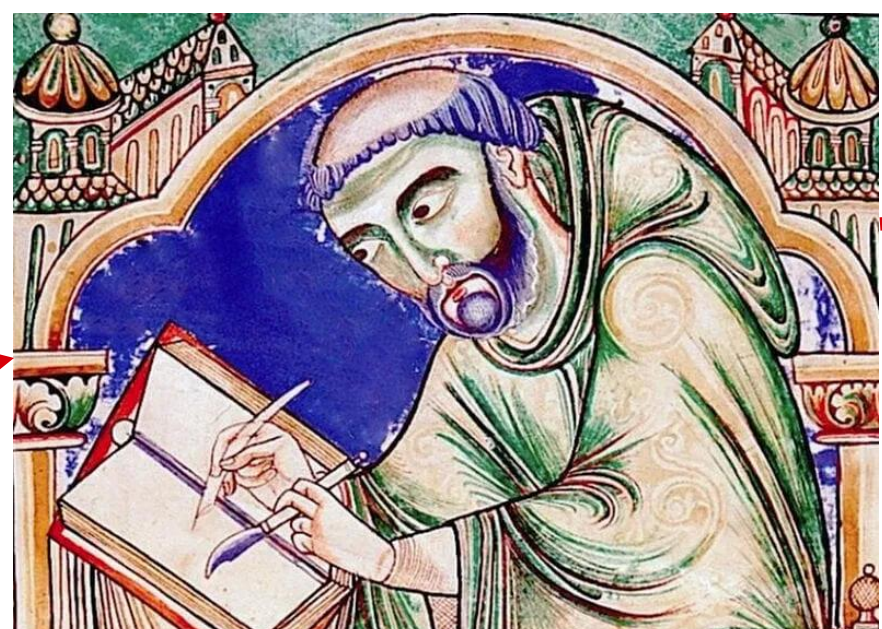
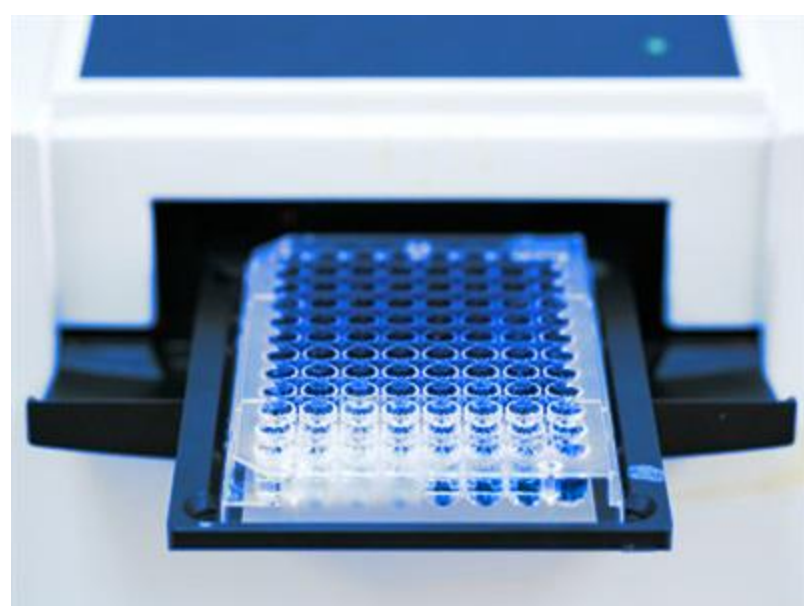
- organism**
 - protozoan ? (1 of 47)
 - Plasmodium falciparum ?
 - Plasmodium falciparum 3D7 ? (2 of 3)
 - prokaryote ?
 - virus ?
 - absence ?
- assay biology component**
 - biological process
 - biological regulation
 - regulation of biological process
 - regulation of cellular process
 - regulation of cell communication
 - regulation of signal transduction
 - regulation of growth
 - regulation of cell growth

- ❖ Can also use annotations like fingerprints

Curation: now vs. then



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



❖ Make the FAIR UI *easier*

❖ Introduce short term payoffs

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 browser address bar shows the URL: https://app.collaborativedrug.com/vaults/4126/molecules?per_page=50. The page header includes the CDD VAULT logo, the title "Probes Vault", and navigation links for "Help" and "Log out". Below the header, there are tabs for "Explore Data", "Import Data", "Reports", "Settings", and a user profile for "Alex Clark".

The main content area features a search bar with the text "Search molecules..." and a dropdown menu set to "Any field". A "Go" button is next to the search bar. Below the search bar, there are several tabs: "Projects", "Saved Searches", "Collections", "Molecules" (selected), "Protocols", and "Plates".

The "Molecules" tab shows a list of 2815 molecules, with "Showing 50 per page" and "Page 1 of 57". The list includes chemical structures and associated data for several molecules:

- JHU- 12792**: CDD-2655024, Synonyms: (no synonyms), Protocols: 5
- JHU- 12791**: CDD-2654766, Synonyms: (no synonyms), Protocols: 5
- JHU- 12790**: CDD-2626575, Synonyms: (no synonyms), Protocols: 5
- JHU- 12789**: CDD-2654327, Synonyms: (no synonyms), Protocols: 5
- JHU- 12788**: CDD-2655031, Synonyms: (no synonyms), Protocols: 5
- JHU- 12787**: CDD-2654491, Synonyms: (no synonyms), Protocols: 5

The interface also includes a sidebar on the left with a "Create a new ..." button and a list of projects, each with a count in a dropdown menu. The projects listed are: Projects (0), ADMEdata.com for Ekins Ca... (0), Alex Clark Sandbox (0), AZ Public ChEMBL Data (-), 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).

Protocol forms



- ❖ Mix text, numbers & pick lists with selected ontology terms

Edit Form: In-vitro Mouse Cancer Screening

Protocol Definition | Run Definition

Form Name*
In-vitro Mouse Cancer Screening

Protocol Name*

Category*

Description

Instrument

bioassay type [Annotation](#) [Free Text](#)

assay format [Annotation](#) [Free Text](#)

organism [Annotation](#) [Free Text](#)

applies to disease [Annotation](#) [Free Text](#)

physical detection method [Annotation](#) [Free Text](#)

Delete Cancel

Run Data | **Protocol Details** | Projects 1 | Files 0

Form

Protocol Name*
[Special characters](#)

Category*

Description

[Special characters](#)

Instrument

bioassay type [Annotation](#) [Free Text](#) **assay format** [Annotation](#) [Free Text](#)

organism [Annotation](#) [Free Text](#)

applies to disease [Annotation](#) [Free Text](#)

physical detection method [Annotation](#) [Free Text](#)

CANCEL

Business logic

- ❖ Reduce curation efforts
- ❖ *and* fewer mistakes



Pick Term: organism

organism

- cellular organisms
- eukaryote
- + fungi
- metazoa
- + invertebrate
- vertebrate
- mammalian
 - Bos taurus
 - Canis lupus familiaris
 - Capra hircus
 - Cavia porcellus
 - Cercopithecidae
 - Cervus canadensis nelsoni
 - Chinchilla lanigera
 - Chlorocebus aethiops
 - Chlorocebus aethiops
 - Cricetinae
 - Equus caballus
 - Felis catus
 - Homo sapiens
 - marmosets
 - Marmota monax
 - Mastomys coucha
 - Meriones unguiculatus
 - Mesocricetus auratus
 - Microtus ochrogaster
 - Murinae
 - Mus musculus**
 - Mustela putorius furo
 - Oryctolagus cuniculus
 - Ovis aries
 - Pan troglodytes
 - Papio anubis
 - Papio hamadryas
 - Potorous tridactylus

Pick Term: bioassay type

bioassay type

- ADMET
- binding type
- **functional**
- functional phenotypic
- functional target-based
- physicochemical

Pick Term: applies to disease

Disease

- + disease by infectious agent
- + disease of anatomical entity
- disease of cellular proliferation
- + benign neoplasm
- cancer
 - + cell type cancer
 - head and neck cancer
 - organ system cancer
 - + cardiovascular cancer
 - endocrine gland cancer
 - + neuroendocrine tumor
 - **pancreatic cancer**
 - islet cell tumor
 - insulinoma
 - pancreatic endocrine carcinoma
 - pancreatic gastrinoma
 - pancreatic somatostatinoma
 - malignant exocrine pancreas neoplasm
 - pancreatic carcinoma
 - pancreatic ductal carcinoma
 - pancreatoblastoma
 - + pineal gland cancer
 - + pituitary cancer
 - + thymus cancer
 - + peritoneum cancer
 - + reproductive organ cancer
 - + retroperitoneal cancer
 - + thoracic cancer
 - + urinary system cancer
- + pre-malignant neoplasm
- + disease of mental health
- + disease of metabolism
- + genetic disease
- + physical disorder
- + syndrome

bioassay type ⊕ Annotation ⊕ Free Text ✎ assay format ⊕ Annotation ⊕ Free Text ✎

organism ⊕ Annotation ⊕ Free Text ✎

applies to disease ⊕ Annotation ⊕ Free Text ✎

physical detection method ⊕ Annotation ⊕ Free Text ✎

Pick Term: physical detection method

physical detection method

- + fluorescence method
- + imaging method
- + label free method
- luminescence method
 - alphascreen
 - bioluminescence
 - bioluminescence resonance energy transfer
 - **chemiluminescence**
 - chemiluminescence resonance energy transfer
- mass spectrometry
- + radiometry method

default

subset

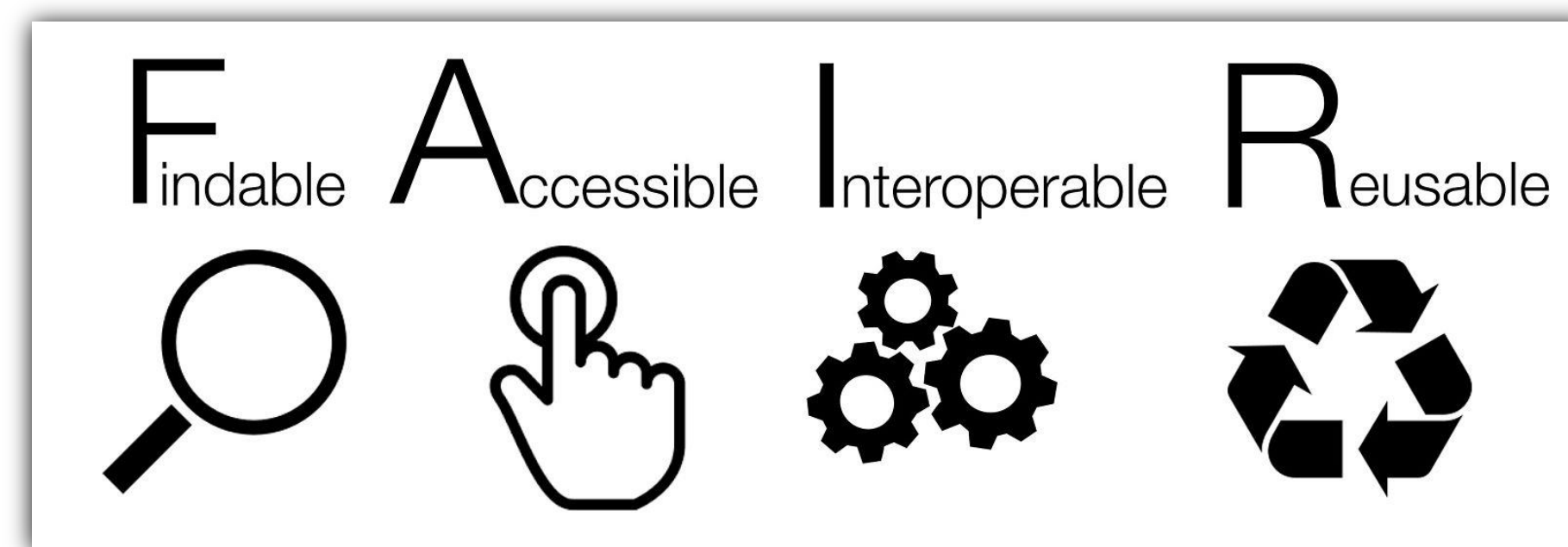
locked

dependence

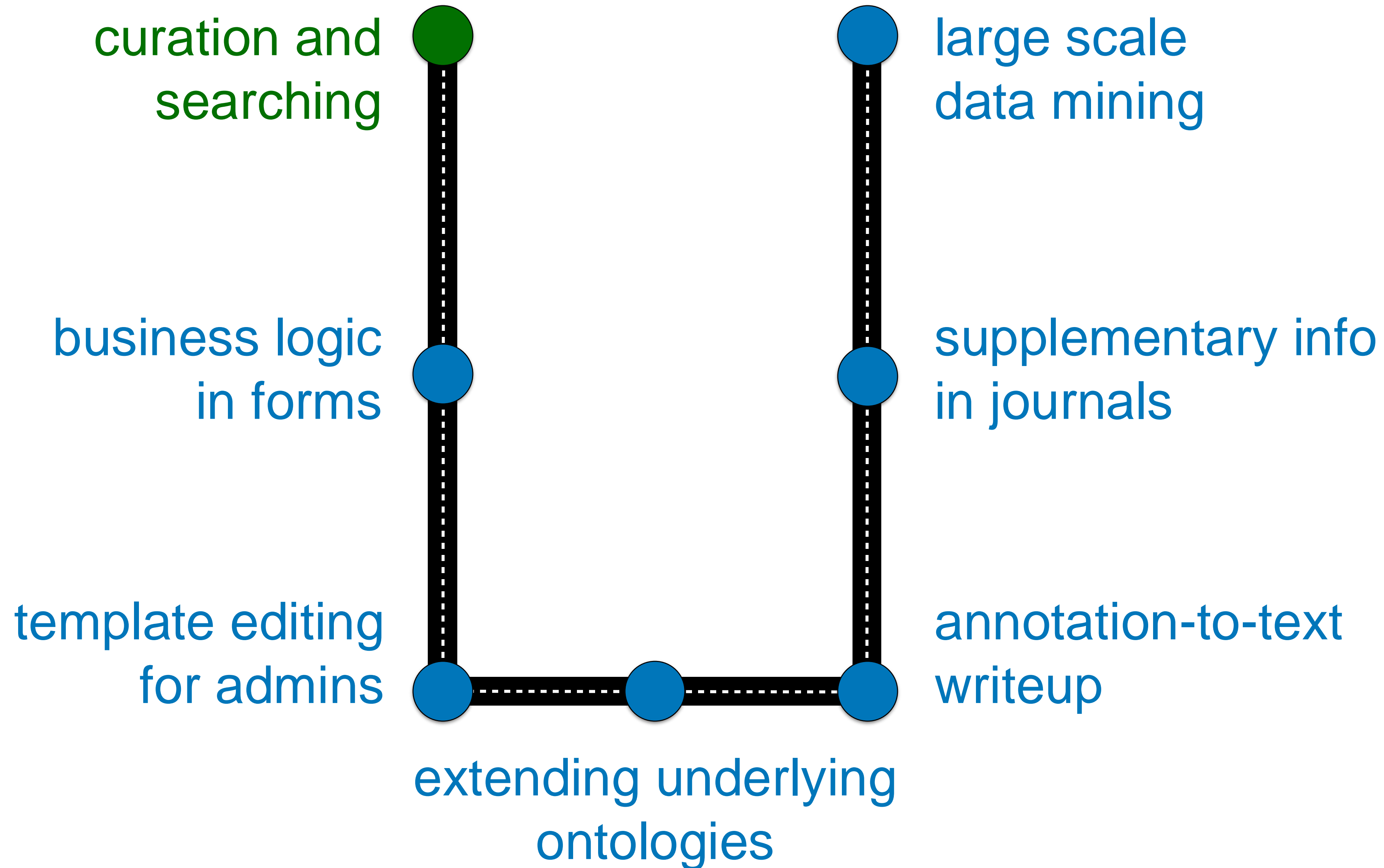
Collaboration



- ❖ 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