



CEDAR

CENTER FOR EXPANDED DATA ANNOTATION
AND RETRIEVAL



metadatacenter.org

(2014 -2018)

Steering Committee



Mark Musen,
Principal Investigator

Steven Kleinstein

Susanna-A Sansone

Jeffrey Wiser

Carol Bean,
Project Manager

Kei-Hoi Cheung

Michel Dumontier

Olivier Gevaert

Purvesh Khatri

Community-developed *content standards*

- To ***structure, enrich*** and ***report*** the description of the datasets and the experimental context under which they were produced
- To facilitate **discovery, sharing, understanding** and **reuse** of datasets



formats



Including ***conceptual model, conceptual schema*** from which an exchange format is derived **to allow data to flow from one system to another**

terminologies



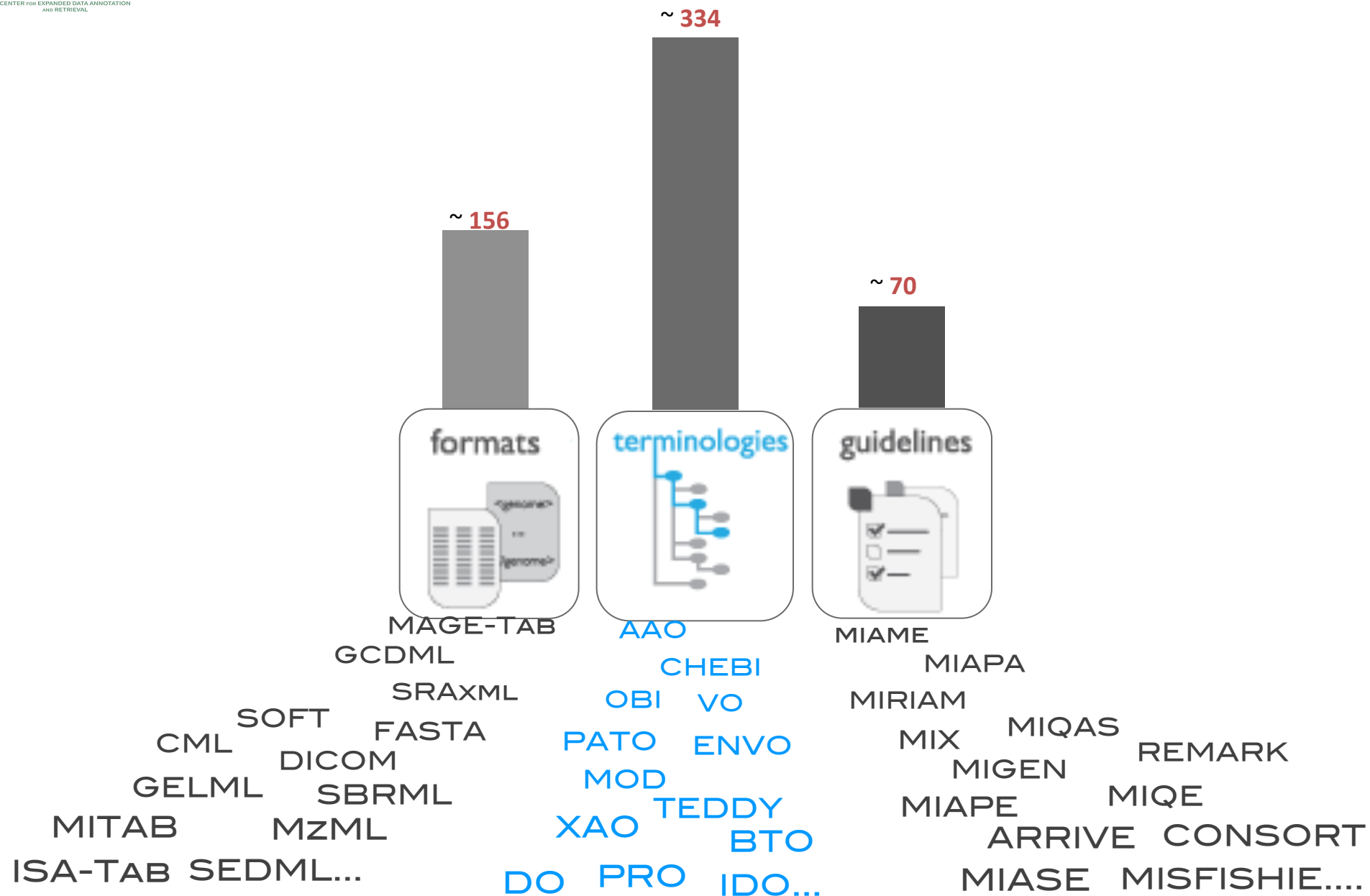
Including ***controlled vocabularies, taxonomies, thesauri, ontologies*** etc. **to use the same word and refer to the same 'thing'**

guidelines



Including ***minimum information reporting requirements, or checklists*** **to report the same core, essential information**

In the life sciences alone.....almost 600!



Researchers *hate* standards!

- Most researchers understand the value of standardized descriptions, when using third-party datasets
- But when asked to structure their datasets, they view requests for even “minimal” information as burdensome

HOME SEARCH SITE MAP

NCBI > GEO > **Accession Display** ?

GEO help: Mouse over screen elements for information.

Scope: Format: Amount:

Series GSE35240

Status	Public on Aug 20, 2012
Title	Gene expression in mitotic tissues with too many centrosomes
Organism	Drosophila melanogaster
Experiment type	Expression profiling by array
Summary	Centrosome defects are a common feature of many cancers and can proceed through the major pathways of centrosome amplification. In most of these cases, centrosome defects do not cause cell death because they can adapt to cope with any number of centrosomes and centrosome amplification prevents cells from assessing how centrosome loss or over-replication affects cell growth by profiling the global transcriptome. We have used this approach to assess expression of the mutant strains that either lack centrosomes or have too many centrosomes.
Overall design	Mitotic tissues (brains and imaginal discs) from <i>Drosophila</i> larvae of mutants lacking centrosomes (Sally and OregonR). We extracted RNA from these tissues and used it for hybridisation to Affymetrix arrays. The biological sample, material and methods used for expression of the mutant strains are described in the accompanying paper.
Contributor(s)	Baumbach J , Levesque MP , Raff J
Citation(s)	Baumbach J, Levesque MP, Raff J. Centrosome defects dramatically perturb global gene expression in <i>Drosophila</i> . <i>Development</i> 15;1(10):983-93. PMID: 23213333

Researchers *hate* standards!

- Most researchers understand the value of standardized descriptions, when using third-party datasets
- But when asked to structure their datasets, they view requests for even “minimal” information as burdensome
- There is an urgent need to lower the bar for authoring good metadata

HOME SEARCH SITE MAP

NCBI > GEO > **Accession Display** ?


GEO help: Mouse over screen elements for information.

Scope: Format: Amount:

Series GSE35240

Status	Public on Aug 20, 2012
Title	Gene expression in mitotic tissues with too many centrosomes
Organism	Drosophila melanogaster
Experiment type	Expression profiling by array
Summary	Centrosome defects are a common feature of many cancers and can proceed through the major pathways of centrosome amplification. In most of these cases, centrosome defects do not cause cell death because they can adapt to cope with any number of centrosomes and centrosome amplification prevents cells from assessing how centrosome loss or over-replication affects cell growth by profiling the global transcriptome. We used Affymetrix arrays to assess expression of the mutant strains that either lack centrosomes or have too many centrosomes.
Overall design	Mitotic tissues (brains and imaginal discs) from <i>Drosophila</i> larvae of mutants lacking centrosomes (Savage and OregonR). We extracted RNA from these tissues and used it for hybridisation to Affymetrix arrays. We used the biological sample, material description and expression of the mutant strains.
Contributor(s)	Baumbach J , Levesque MP , Raff J
Citation(s)	Baumbach J, Levesque MP, Raff J. Centrosome defects dramatically perturb global gene expression in <i>Drosophila</i> . <i>Genes Dev</i> 15;1(10):983-93. PMID: 232133

Our two initial uses cases



IMMPORT

BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY


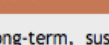
[Home](#)
[About ImmPort](#)
[Access Data](#)
[Tools](#)
[Resources](#)
[News & Events](#)

ImmPort's Mission


ImmPort is a long-term, sustainable data warehouse for the purpose of promoting re-use of immunological data generated by NIAID DAIT and DMID funded investigators.


ImmPort supports analysis of flow cytometry results and HLA genetic associations.

What is ImmPort

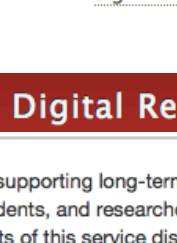



Human Immunology Project Consortium

 **STANFORD UNIVERSITY LIBRARIES**
[Sign in via WebAuth](#) | [Contact Us](#)

 **Stanford Digital Repository – Online Deposit**

The SDR is a service supporting long-term management of scholarly information resources at Stanford. Faculty, students, and researchers use the SDR to promote and protect the products of their work. The benefits of this service distinguish the SDR from other content storage or management options on campus: deposited scholarly content is preserved in a robust, reliable, and secure environment and is available from persistent URLs (PURLs) with optional access controls.



```
graph LR; You[You] --> SL[Stanford Libraries]; SL --> WU[Web Users];
```

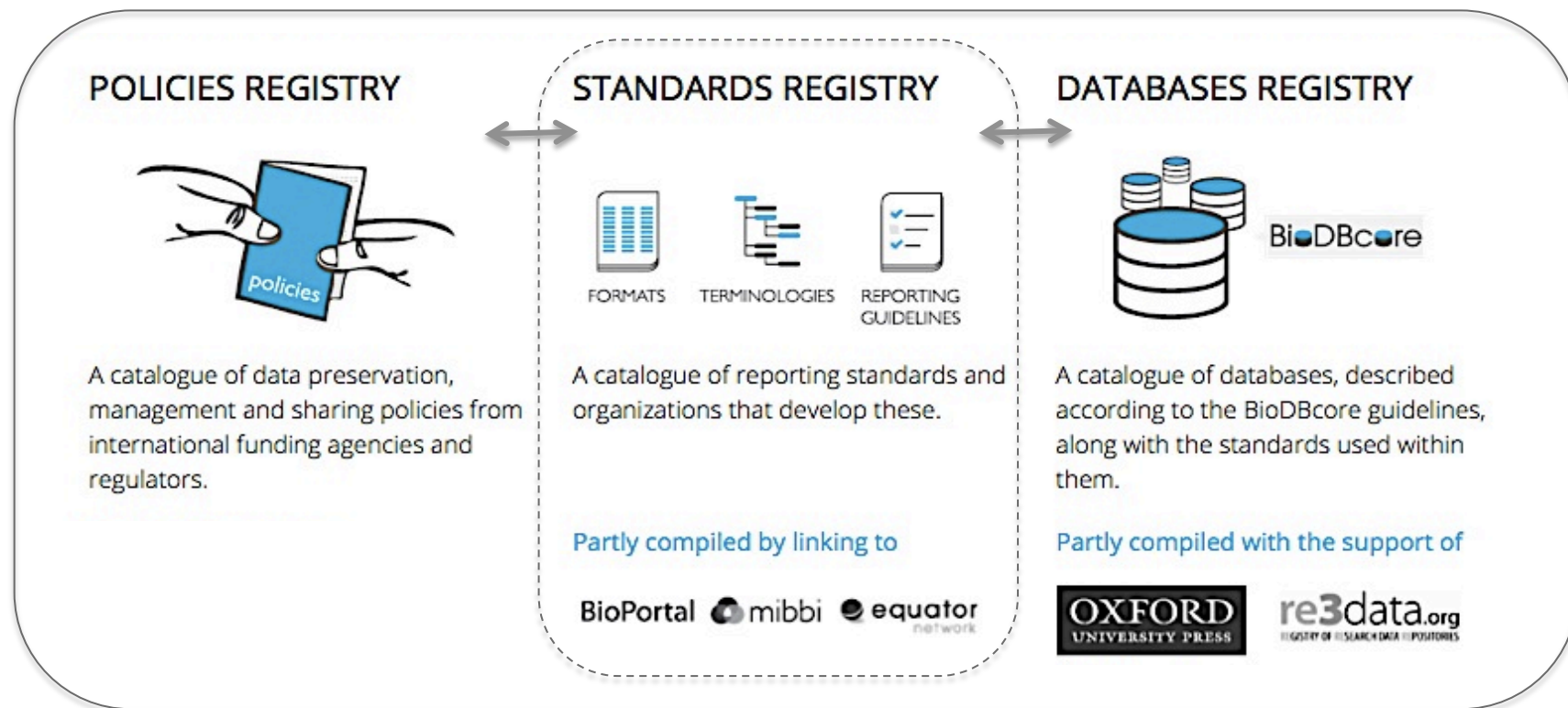
The diagram illustrates the workflow of the Stanford Digital Repository (SDR) service. It consists of three main components: **You**, **Stanford Libraries**, and **Web Users**.

- You**: Represented by a person icon. Below the icon, it says "...deposit items".
- Stanford Libraries**: Represented by a database icon. Below the icon, it says "...provides long-term preservation".
- Web Users**: Represented by three person icons. Below the icons, it says "...discover your deposits via SearchWorks". Above the icons, it says "...access your deposits via PURLs".

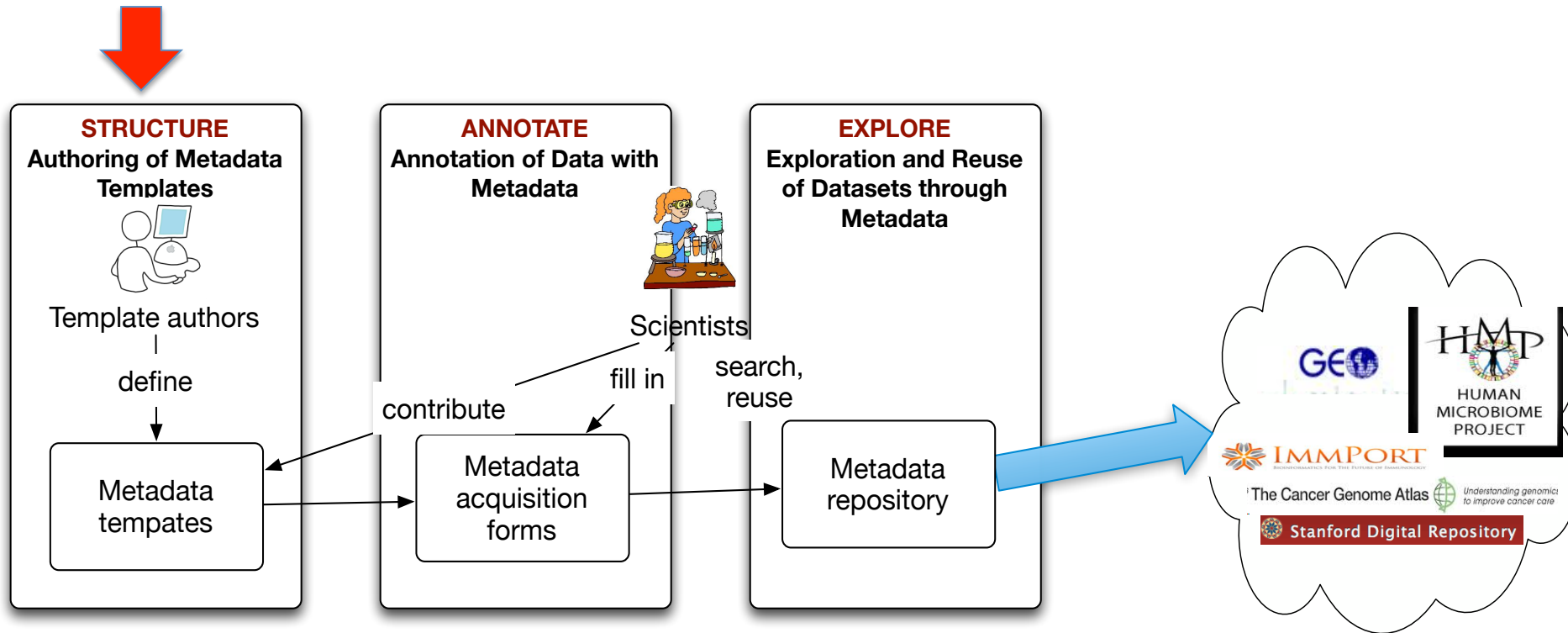
Arrows indicate the flow of the process: from **You** to **Stanford Libraries**, and from **Stanford Libraries** to **Web Users**.

1. Map the landscape of content standards

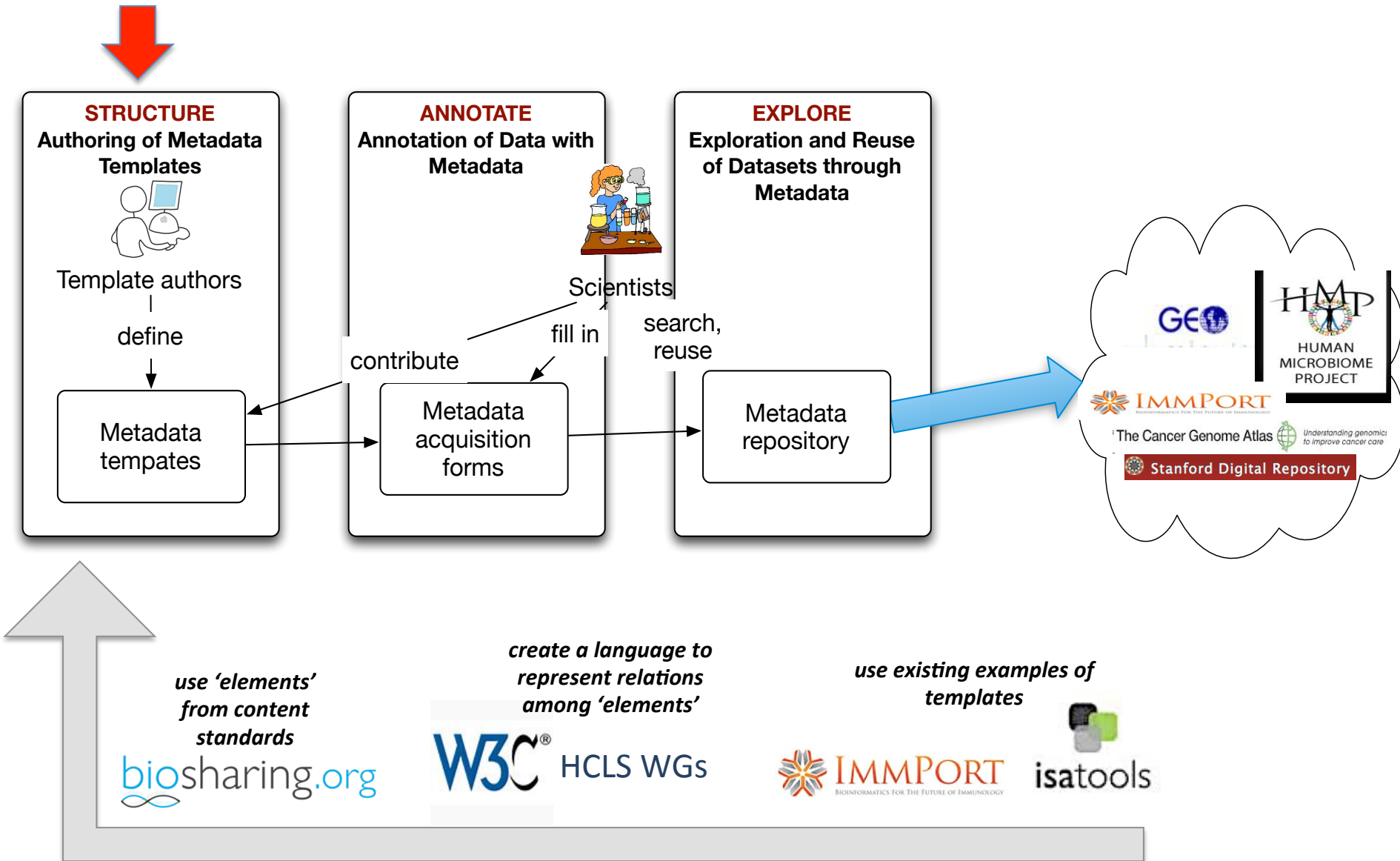
biosharing.org



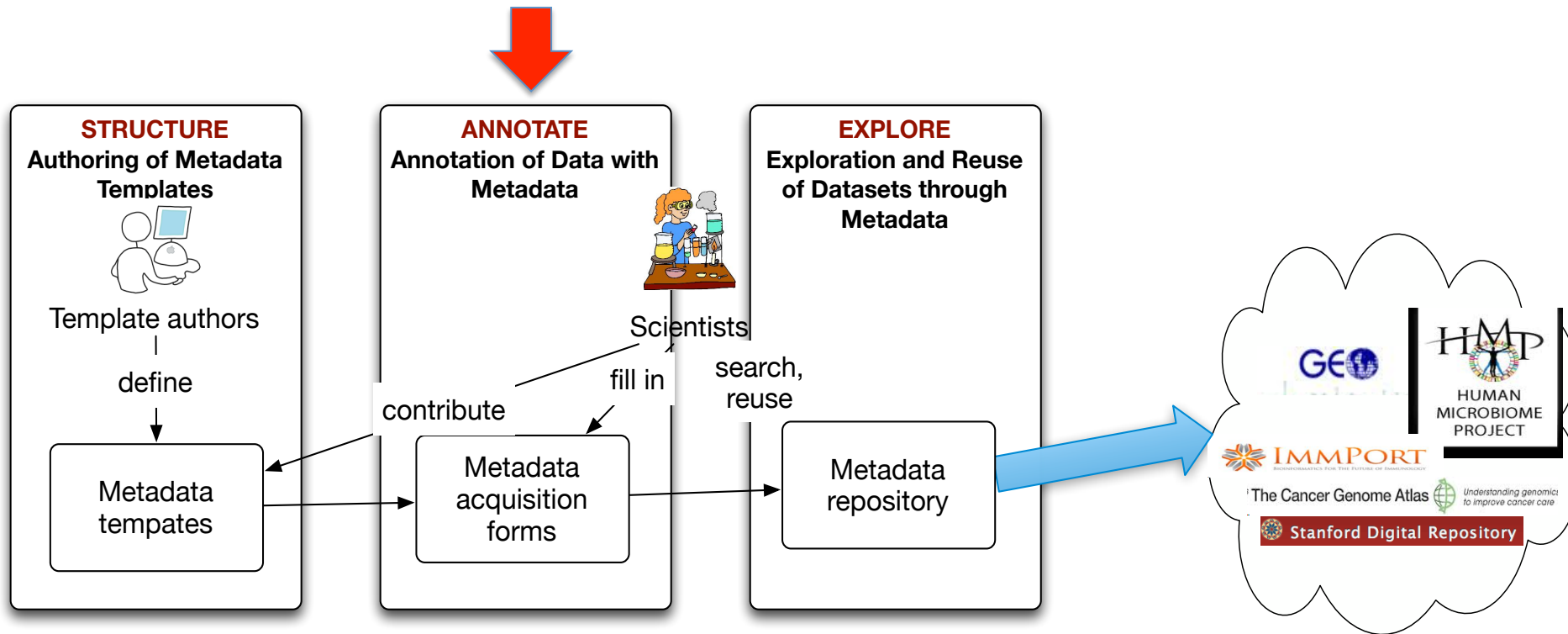
2. Develop methods for creating templates



2. Develop methods for creating templates

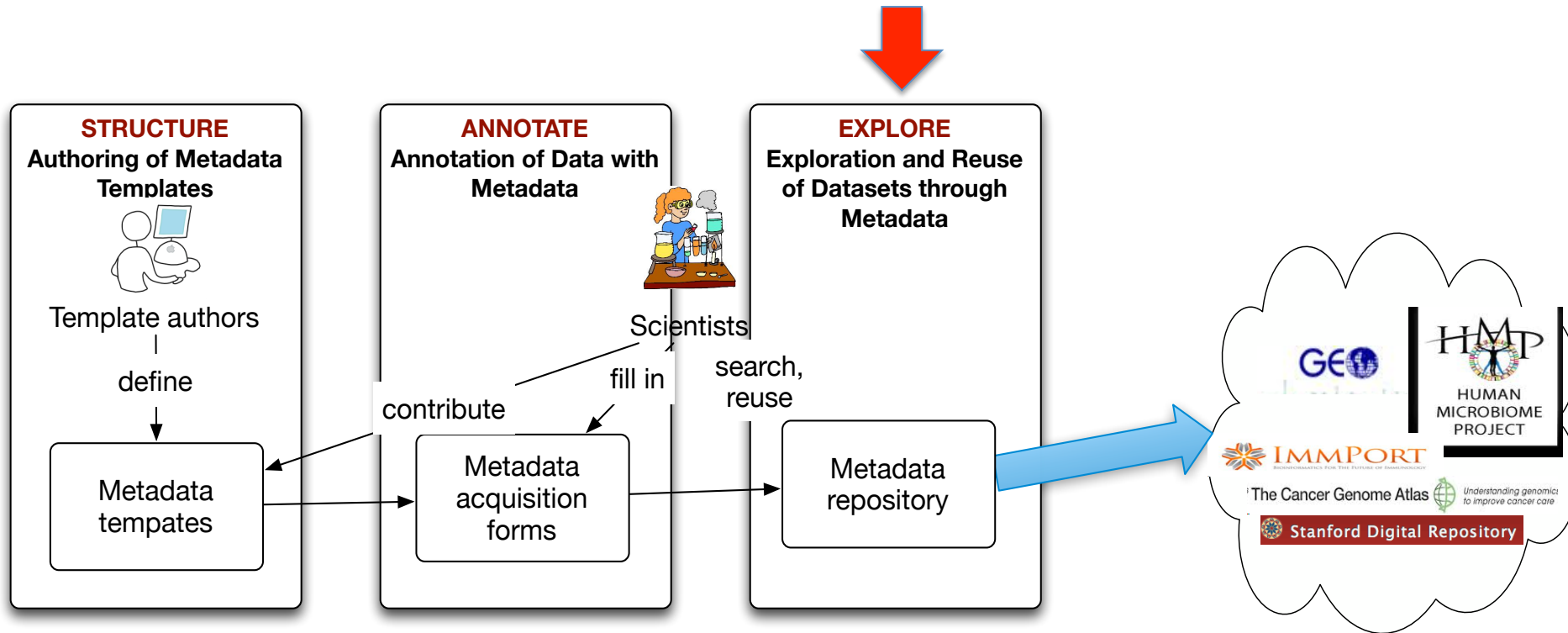


3. Develop methods to ease use of templates



- Enable researchers to help us creating templates appropriate to their needs
- Help researchers to *find* and *use* these templates to describe their experiments, and *populate* them with appropriate values (e.g. terms from ontologies)

4. Create a repository of populated templates



CEDAR repository will:

- **store** the experimental descriptions
- **facilitate** submission of datasets to our two case study repositories and progressively to other recognized online repositories

5. Exploring ways to enhance metadata

- Analyze the CEDAR repository to reveal ***patterns*** in the metadata that will enable the metadata tools to use ***predictive data entry*** to ease the task of filling out the templates
 - Augment those metadata with links to the published literature (including secondary analyses and retractions!)
 - Augment those metadata with links to follow-up experiments (in online databases and in the literature)
 - Allow the scientific community to comment on the experiment through structured metadata
- Learn how to ease the authoring of metadata, using community standards, to enhance the richness of the experimental descriptions



CEDAR

CENTER FOR EXPANDED DATA ANNOTATION
AND RETRIEVAL



metadatacenter.org

(2014 -2018)