Reports of the Death of Ontologies Are Greatly Exaggerated

Mark A. Musen
Stanford Medicine Professor of Biomedical Informatics Research
Division of Computational Medicine
Stanford University School of Medicine
musen@stanford.edu

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

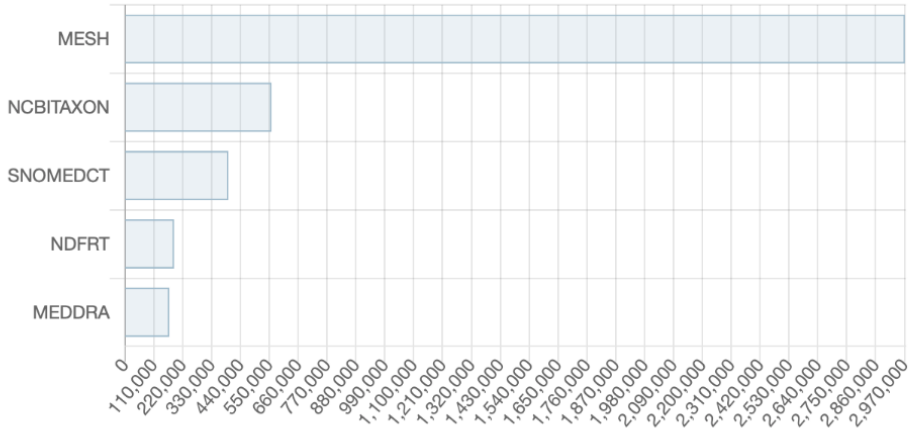
Advanced search

Find an ontology

Start typing ontology name, then choose from list

Browse ontologies

Ontology visits (September 2025)



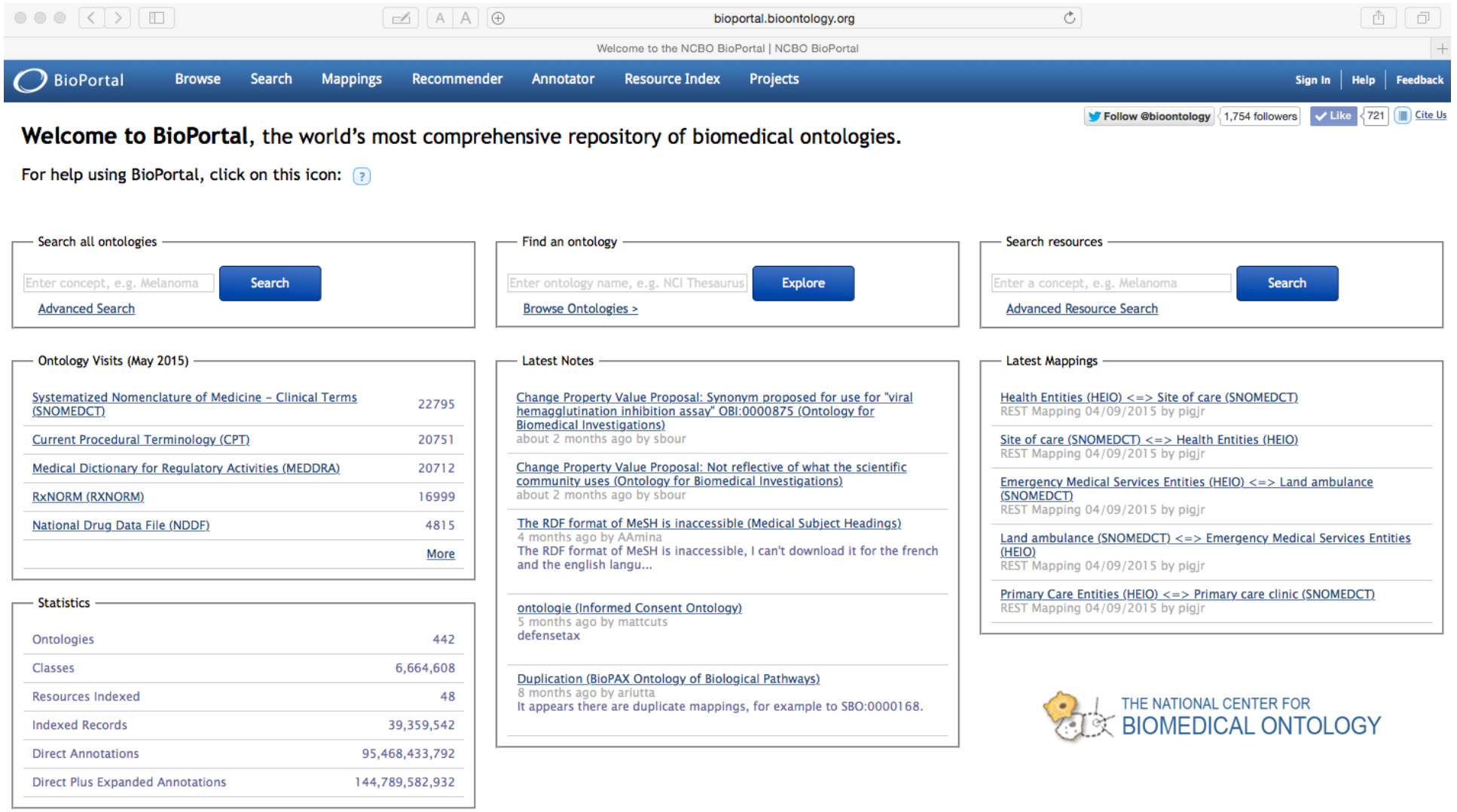
More

Statistics

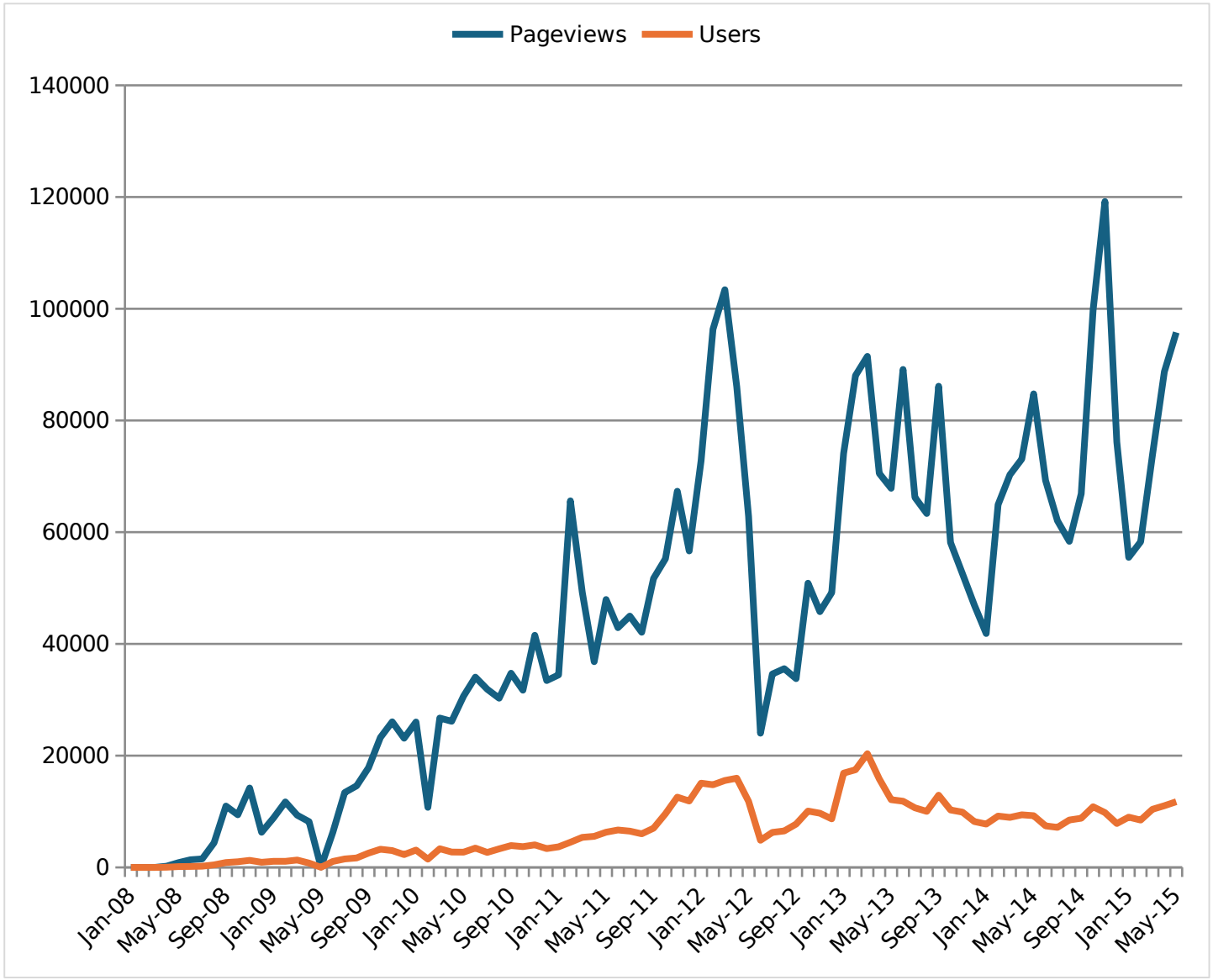
Ontologies	1,231
Classes	17,545,159
Properties	36,286
Mappings	92,327,800

But first, a sad memory

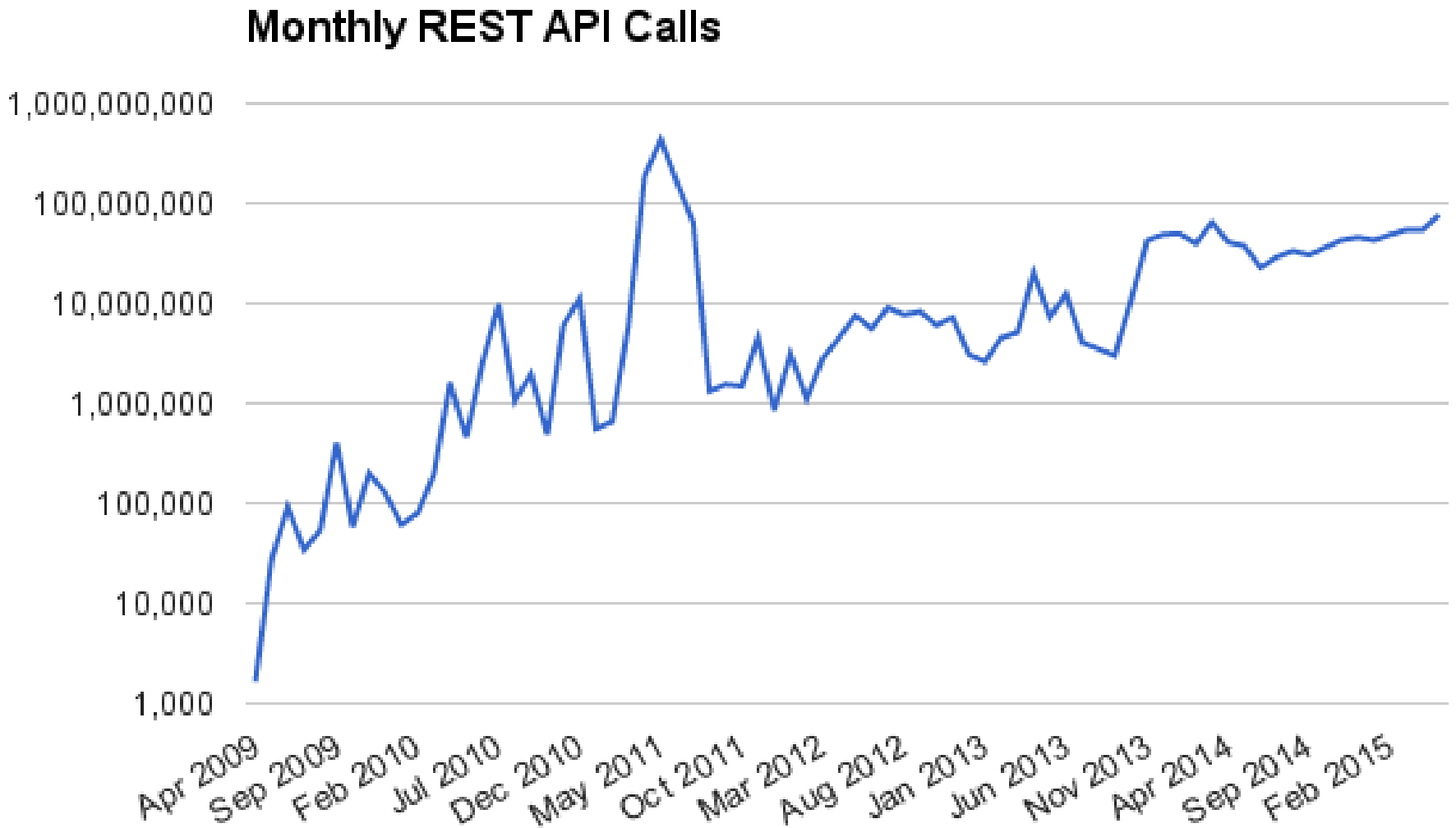
...



NCBO BioPortal Page Views



NCBO BioPortal Internet Traffic



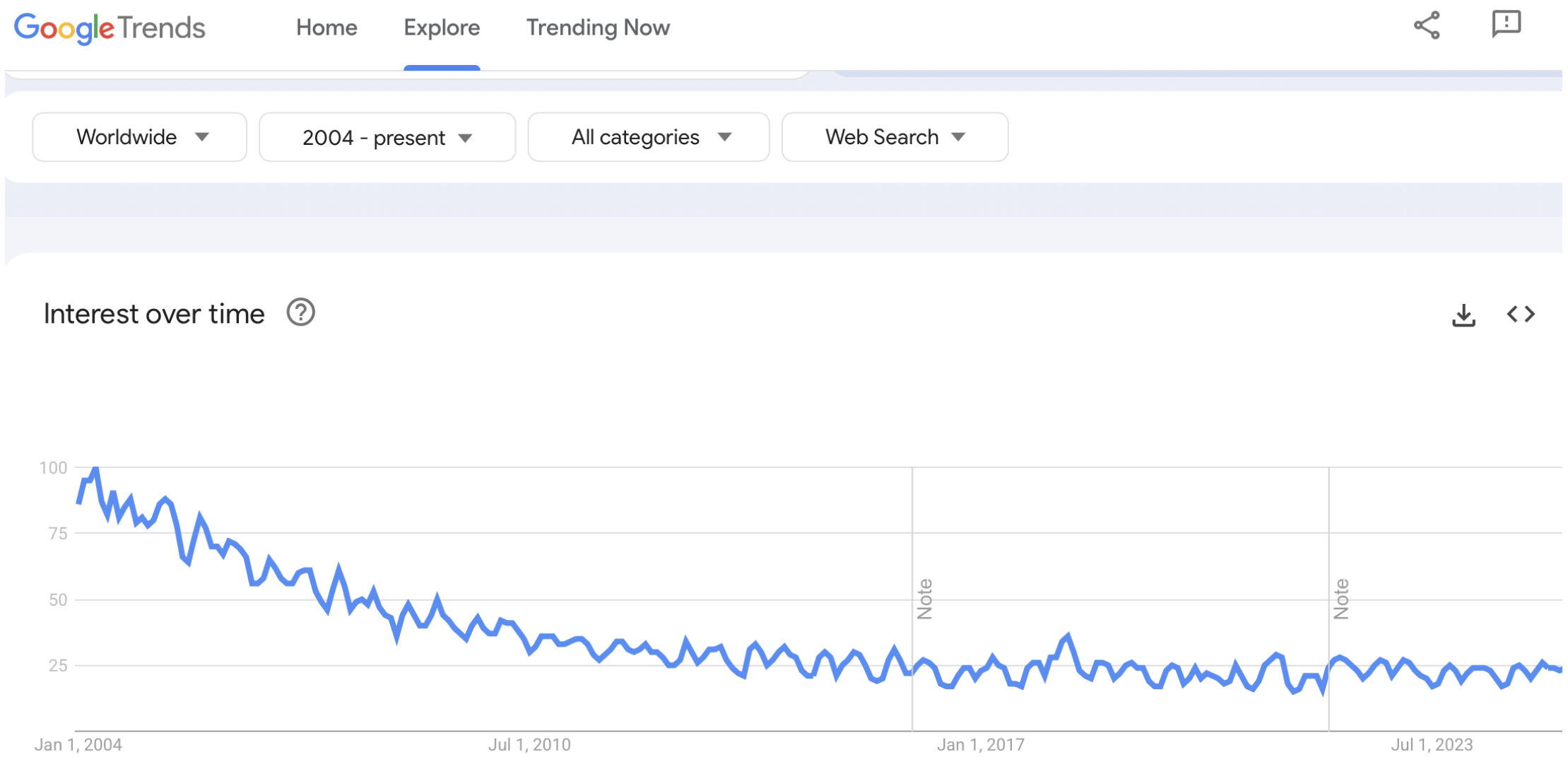
NCBO: Billions and Billions Served ...

- 540,000 unique users
- 4,100,000 page views
- 1,900,000,000 API calls

But then our Scientific Advisory Board said …

- That’s really exciting.
- Now, can you do something *other than* ontologies.

Interest in “ontology” over time



What have I kept hearing since that SAB meeting?

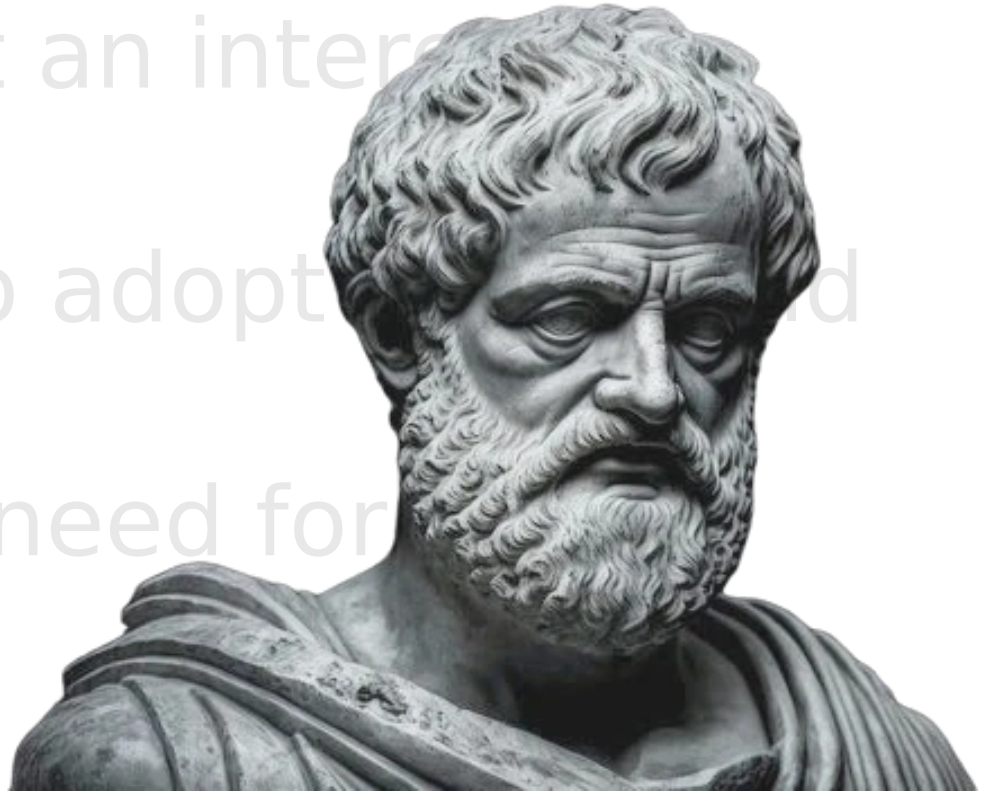
- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt outdated views
- Generative AI obviates the need for ontologies

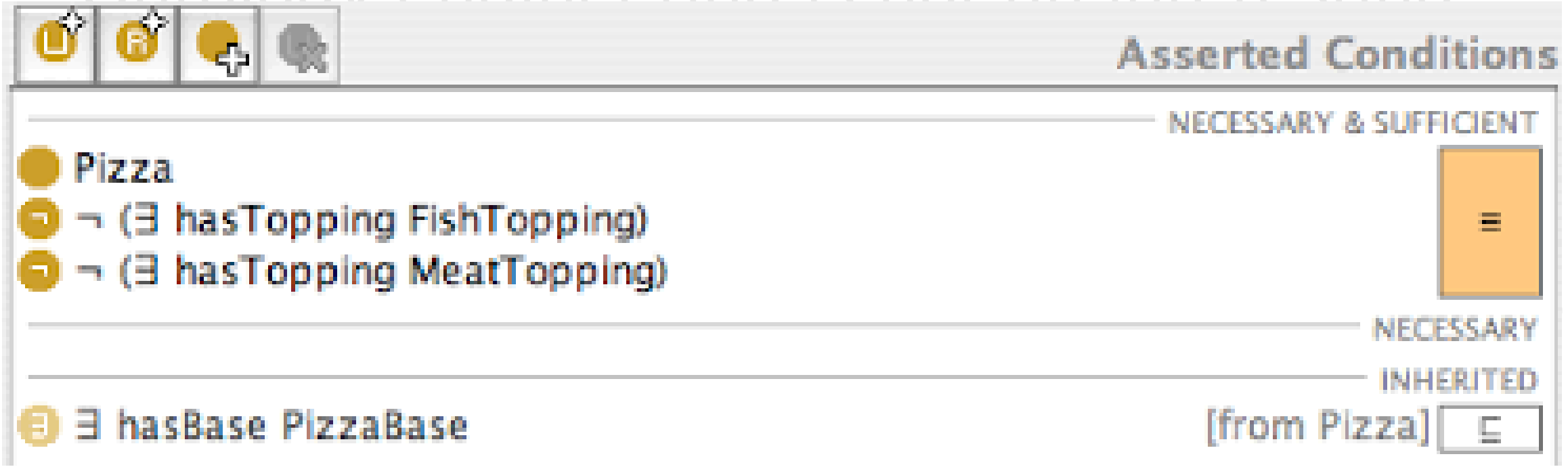


What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting



What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce peo views
- Generative AI obviates ontologies

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

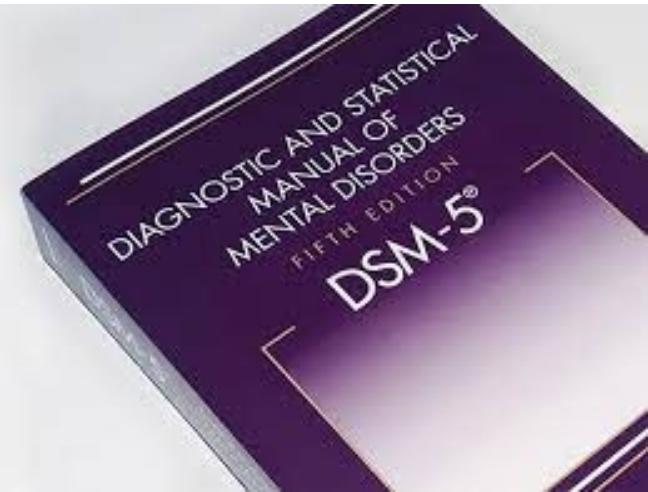
Ontologies in the Behavioral Sciences

Accelerating Research and the Spread of Knowledge

Robert M. Kaplan, Chair
Katy Börner, Committee Member
Bruce F. Chorpita, Committee Member
David Danks, Committee Member
Mark A. Musen, Committee Member
Vimla L. Patel, Committee Member
Timothy J. Strauman, Committee Member
James F. Woodward, Committee Member
Alix Beatty, Study Director

Motivations for the committee

- There has been an explosion of overlapping theory in experimental psychology, mainly because promoting a new theory seems to be required to obtain tenure
- The lack of standard vocabularies makes it impossible to understand how one theory may relate to another
- The lack of standard vocabularies makes it impossible to relate theories of behavior to advances in neuroscience
- Diagnosis in clinical psychology is confused b the lack of a consistent, axiomatized nosolog



Concerns the committee had to deal with

- Ontologies constrain what people can think
- Development of new theories will be difficult if they have to be expressed in terms of existing, approved concepts
- Investigators who ignore ontologies
  - Will be able to be more creative
  - Will achieve greater academic recognition (i.e., tenure)
- Behavioral science is simply too complicated to comport with any ontology or taxonomy

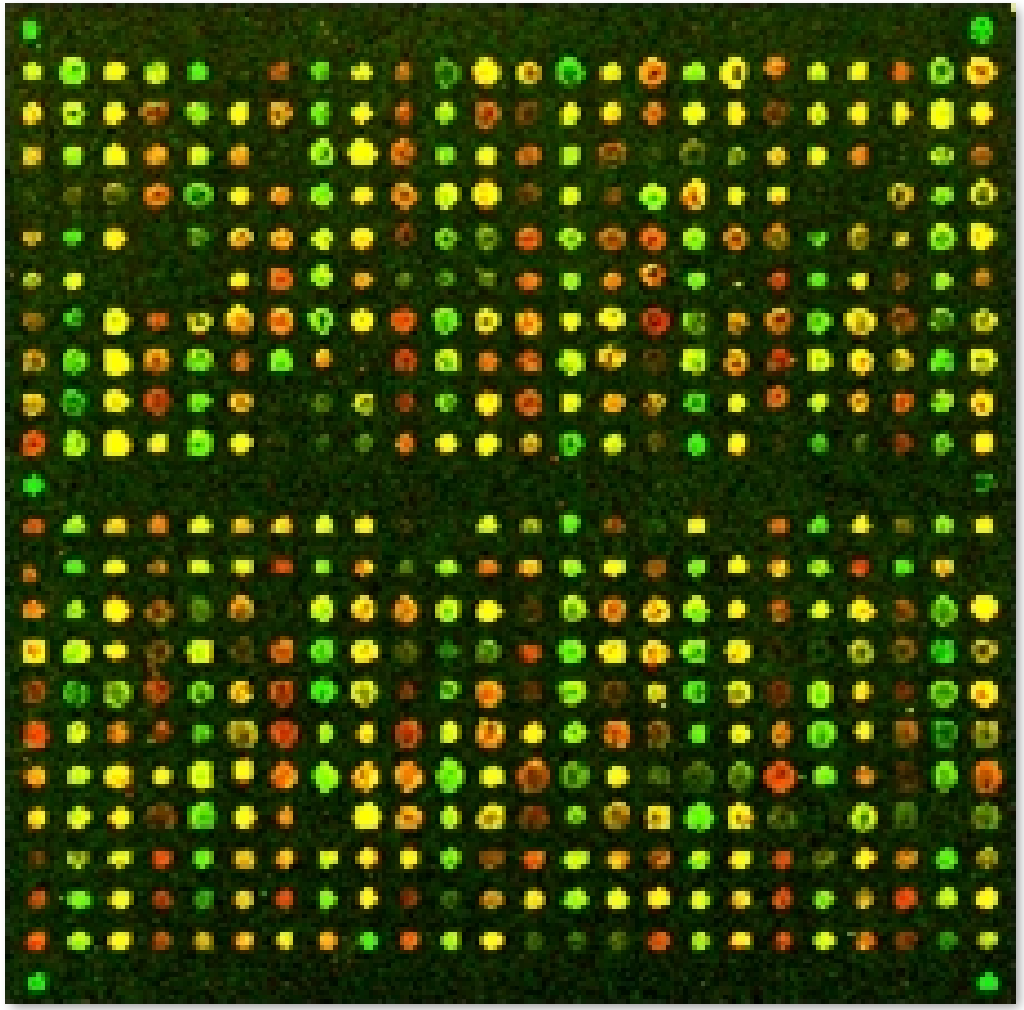
CONCLUSION: Ontology development and use has the potential to move behavioral science forward from a domain in which research is generally siloed and the data and results are often incompatible to one in which the evidence is searchable and more easily integrated and in which computer technology is leveraged in the discovery of new relationships, the development of novel hypotheses, and the identification of knowledge gaps.

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

Experimental data need to be described with metadata

- Who did the experiment?
- What was the subject of the experiment?
- What were the experimental conditions?
- What were the results?



Metadata need ontologies!

age
Age
AGE
`Age
age (after birth)
age (in years)
age (y)
age (year)
age (years)
Age (years)
Age (Years)
age (yr)
age (yr-old)
age (yrs)
Age (yrs)

age [y]
age [year]
age [years]
age in years
age of patient
Age of patient
age of subjects
age(years)
Age(years)
Age(yrs.)
Age, year
age, years
age, yrs
age.year
age_years

Gene Expression Omnibus

Full ▾ Send to: ▾

AtMs-SLE-sle1

Identifiers BioSample: SAMN10417071; Sample name: AtMs-SLE-sle1; SRA: SRS4040527

Organism [Homo sapiens](#) (human)
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

...

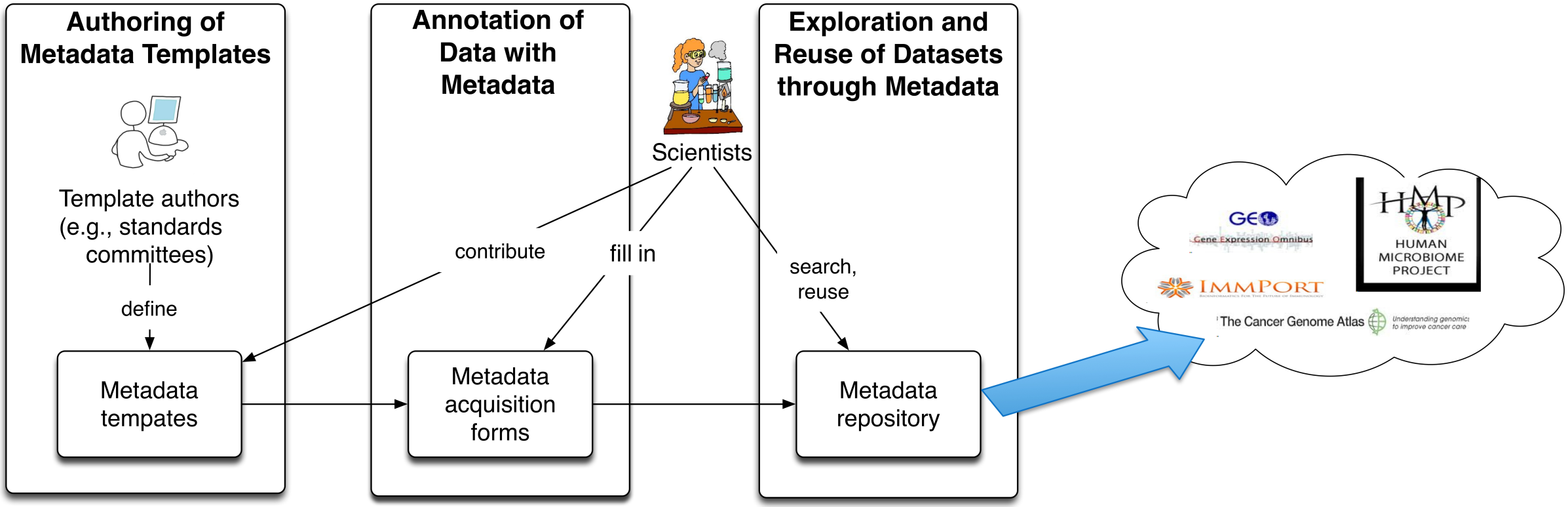
cell subtype	Fresh atypical memory B cells
cell type	Primary cell
disease	SLE
disease stage	New-onset
ethnicity	Asian
health state	SLE
karyotype	46 chromosomes
population	[Peripheral blood](#)
race	yellow race
sample type	leukocyte
treatment	No treatment
IndividaulID	sle1

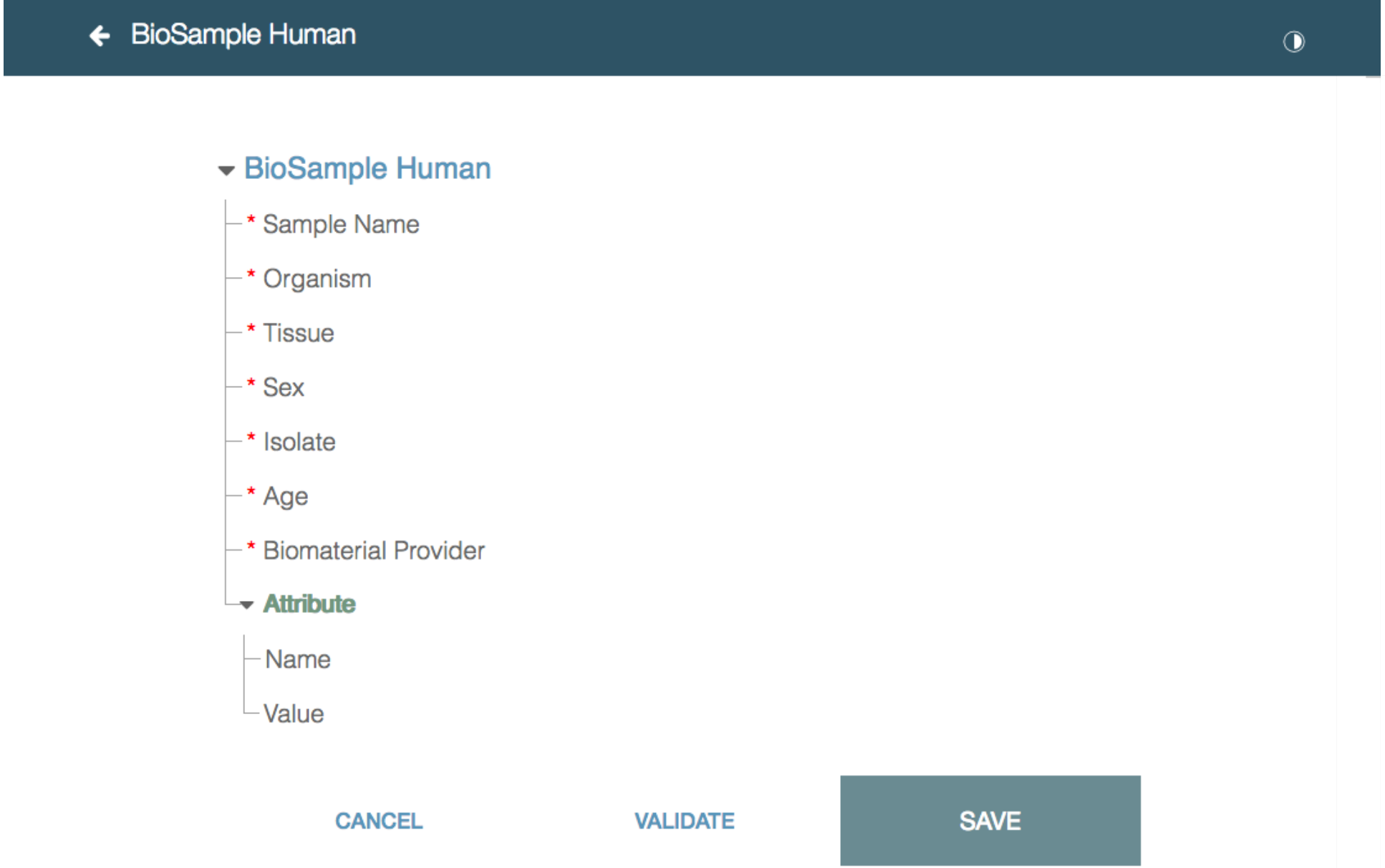
Description Fresh atypical memory B cells from new-onset SLE patient sle1, sorted by Moflo with standard medium RPMI1640

NCBI *BioSample* Metadata are Dreadful!

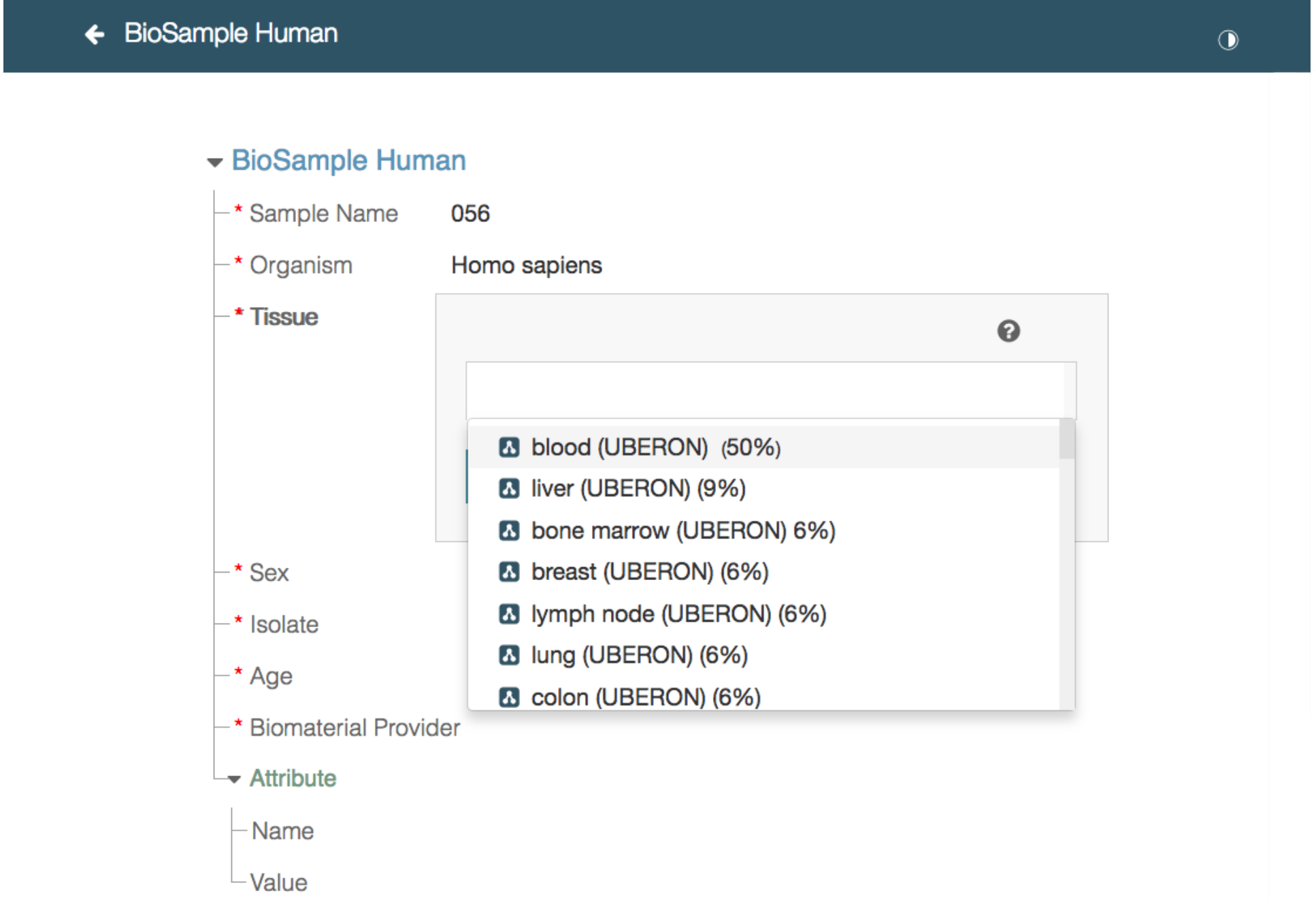
- 73% of “Boolean” metadata values are not actually *true* or *false*
  - *nonsmoker*, *former-smoker*
- 26% of “integer” metadata values cannot be parsed into integers
  - *JM52*, *UVPgt59.4*, *pig*
- 68% of metadata entries that are supposed to represent terms from biomedical ontologies do not actually do so.
  - *presumed normal*, *wild_type*

The CEDAR Workbench





Metadata template for



Metadata template for

Full ▾ Send to: ▾

AtMs-SLE-sle1

Identifiers BioSample: SAMN10417071; Sample name: AtMs-SLE-sle1; SRA: SRS4040527

Organism [Homo sapiens](#) (human)
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

...

cell subtype	Fresh atypical memory B cells
cell type	Primary cell
disease	SLE
disease stage	New-onset
ethnicity	Asian
health state	SLE
karyotype	46 chromosomes
population	[Peripheral blood](#)
race	yellow race
sample type	leukocyte
treatment	No treatment
individualID	sle1

Description Fresh atypical memory B cells from new-onset SLE patient sle1, sorted by Moflo with standard medium RPMI1640

Can't we just fix up the metadata mess with LLMs?

Metadata “correction” with GPT-4

isolate : TN_32

age : 67

biomaterial provider : Prof. Atsushi Kaneda, Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba 260-8670 Japan

sex : female

tissue : lung cancer

Metadata “correction” with GPT-4

isolate : TN_32

age : 67

biomaterial provider : Prof. Atsushi Kaneda, Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba 260-8670 Japan

sex : female

tissue : lung cancer

GPT-4

Sample Isolate Identifier: TN_32

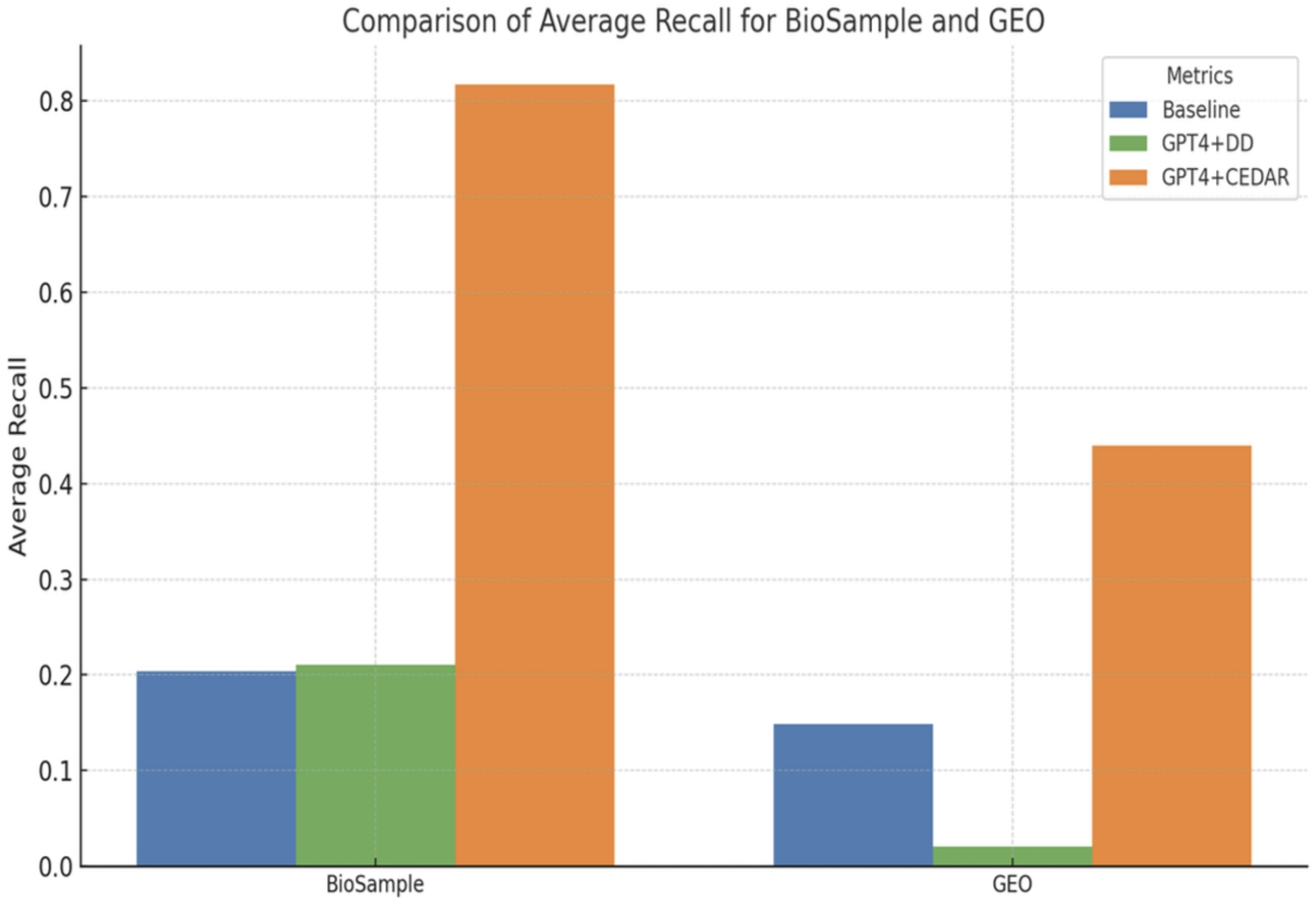
Age of Donor: 67

Sample Contributor: Prof. Atsushi Kaneda, Graduate School of Medicine, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba 260-8670 Japan

Sex of Donor: Female

Sampled Tissue: Lung (afflicted with cancer)

LLMs do a terrible job cleaning up old metadata —until they have access to a CEDAR template!



What can we do to quell these concerns?

- Acknowledge that ontology engineering is hard, but that good tools can make it easier
- Accept that OWL may not be the “one ring to rule them all”
- Educate people about the importance of standards for data management
- Accept that important standards come in forms that are not ontological

What can we do to quell these concerns?

- Acknowledge that ontology engineering is hard, but that good tools can make it easier
- Accept that OWL may not be the “one ring to rule them all”
- Educate people about the importance of standards for data management
- Accept that important standards come in forms that are not ontological

Interest in “Web Ontology Language”

Worldwide | 2004 - present | All categories | Web Search

Interest over time

What can we do to quell these concerns?

- Acknowledge that ontology engineering is hard, but that good tools can make it easier
- Accept that OWL may not be the “one ring to rule them all”
- Educate people about the importance of standards for data management
- Accept that important standards come in forms that are not ontological

Human sample from Homo sapiens

Identifiers: BioSample: SAMN06290438; Sample name: S26; SRA: SRS1954055

Organism: Homo sapiens (human)
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

Package: Human; version 1.0

Attributes

isolate	missing'
age	missing'
biomaterial provider	Ying Hsiu Su, Blumberg Institute
sex	female
tissue	Liver
disease	HCC

BioProject: PRJNA369667
Retrieve all samples from this project

Submission: The Blumberg Institute, Ying-hsiu Su; 2017-02-02

Accession: SAMN06290438 ID: 6290438
BioProject SRA

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

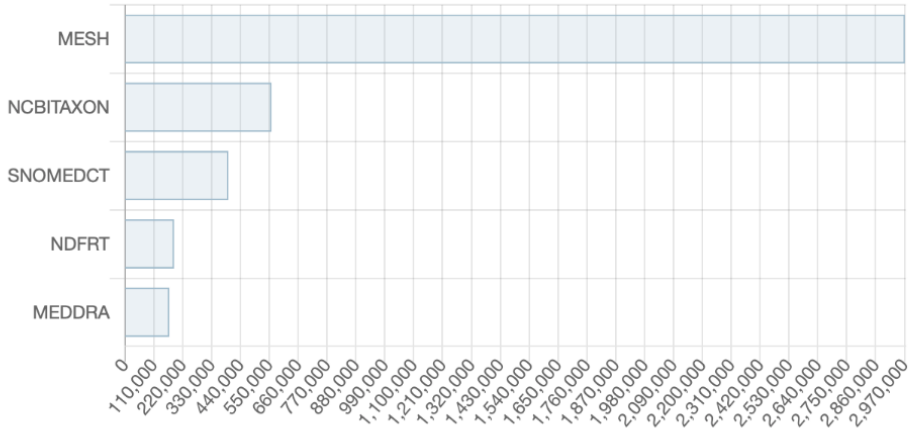
Advanced search

Find an ontology

Start typing ontology name, then choose from list

Browse ontologies

Ontology visits (September 2025)



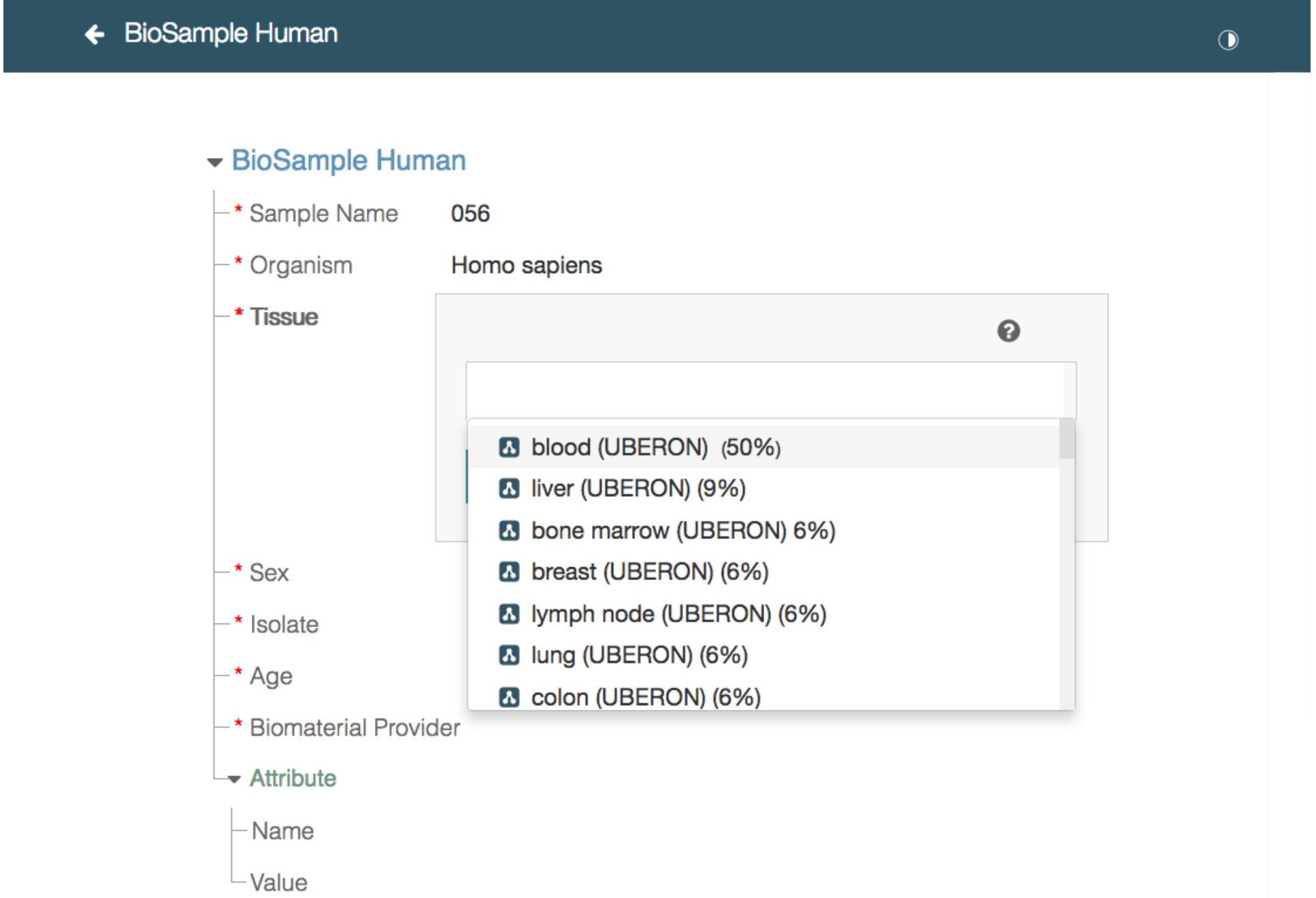
More

Statistics

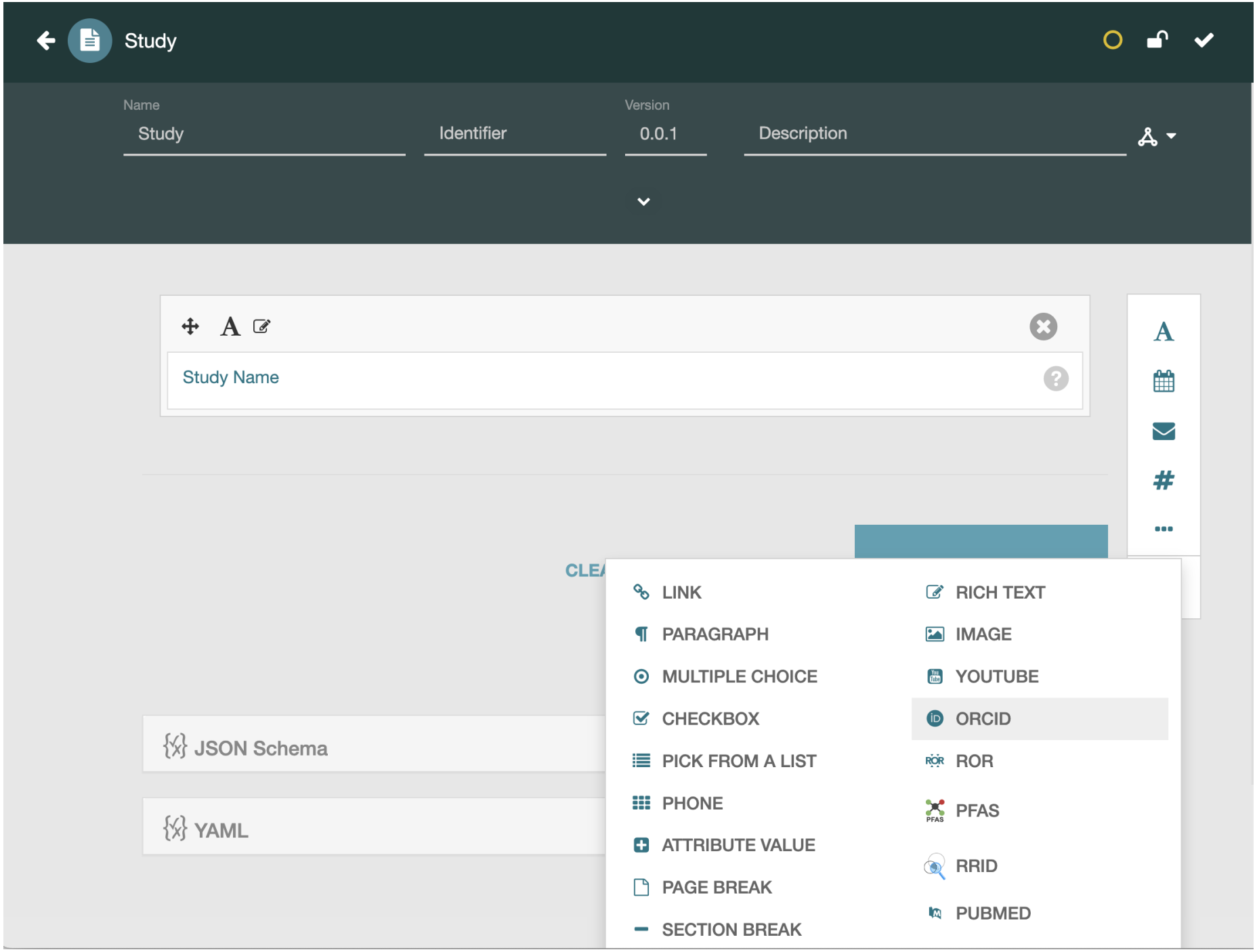
Ontologies	1,231
Classes	17,545,159
Properties	36,286
Mappings	92,327,800

What can we do to quell these concerns?

- Acknowledge that ontology engineering is hard, but that good tools can make it easier
- Accept that OWL may not be the “one ring to rule them all”
- Educate people about the importance of standards for data management
- Accept that important terminological standards come in forms that are not ontological



Metadata template for


Study
Name
Study
Identifier
Version
0.0.1
Description
Study Name
JSON Schema
YAML
LINK
PARAGRAPH
MULTIPLE CHOICE
CHECKBOX
PICK FROM A LIST
PHONE
ATTRIBUTE VALUE
PAGE BREAK
SECTION BREAK
RICH TEXT
IMAGE
YOUTUBE
ORCID
ROR
PFAS
RRID
PUBMED

Study

1.5.0

Expand All Collapse All

Study Name

ROR Institution

Start typing institution name or ROR

iD Principal Investigator

Start typing researcher name or ORCID

Publications (1 .. ∞)

Start typing publication name or PubMed ID

Study

Study

1.5.0

Expand All Collapse All

Study Name

Metadata Powerwash

ROR Institution

stanford

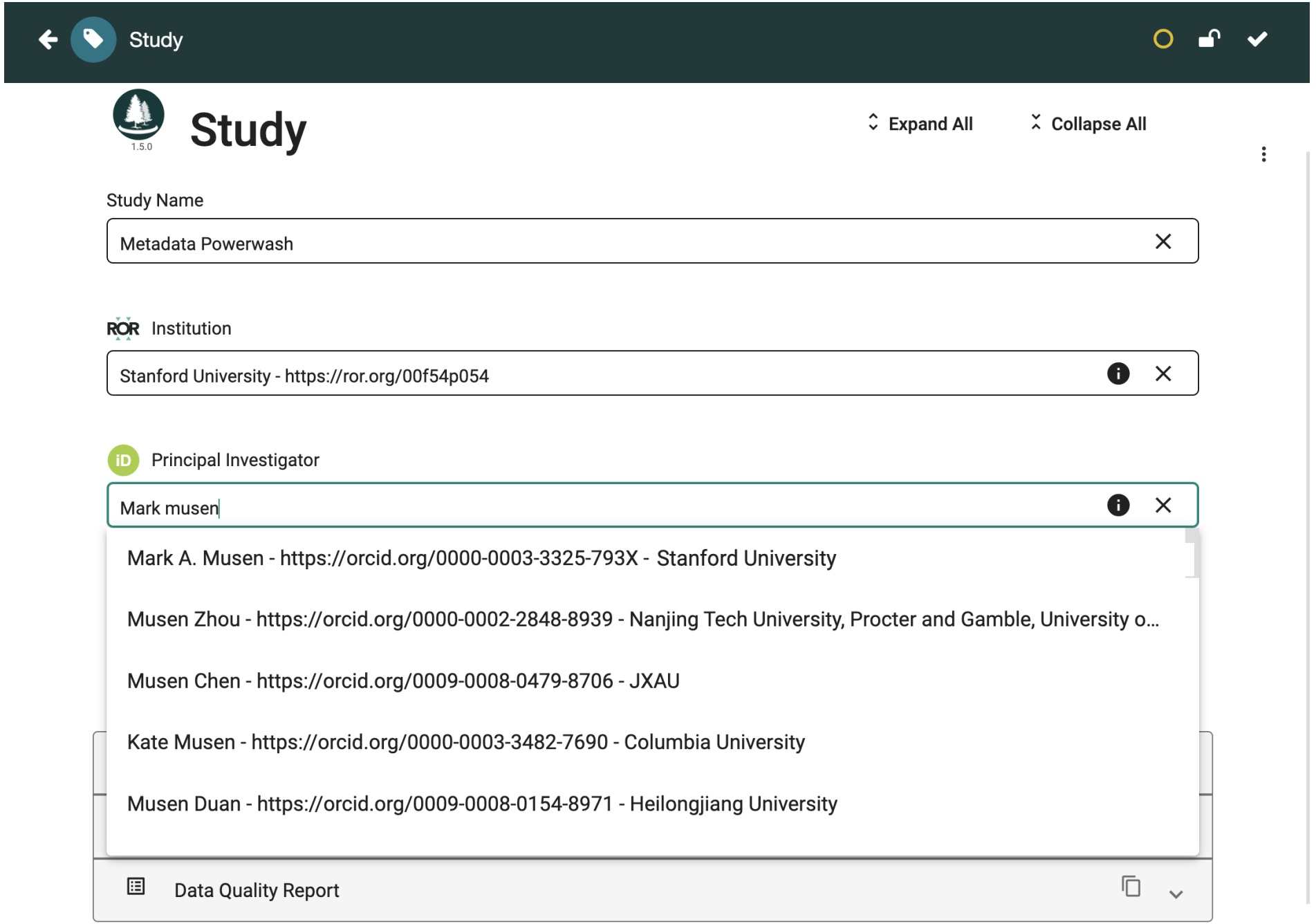
Stanford SystemX Alliance - https://ror.org/0551gkb08

Stanford Cancer Institute - https://ror.org/014qe3j22

Stanford Medicine - https://ror.org/03mtd9a03

Stanford University - https://ror.org/00f54p054

Stanford Health Care - https://ror.org/019wqcg20


Study
Study
1.5.0
Expand All
Collapse All
Study Name
Metadata Powerwash
ROR Institution
Stanford University - https://ror.org/00f54p054
iD Principal Investigator
Mark musen
Mark A. Musen - https://orcid.org/0000-0003-3325-793X - Stanford University
Musen Zhou - https://orcid.org/0000-0002-2848-8939 - Nanjing Tech University, Procter and Gamble, University o...
Musen Chen - https://orcid.org/0009-0008-0479-8706 - JXAU
Kate Musen - https://orcid.org/0000-0003-3482-7690 - Columbia University
Musen Duan - https://orcid.org/0009-0008-0154-8971 - Heilongjiang University
Data Quality Report

Study

1.5.0

Expand All

Collapse All

Study Name

Metadata Powerwash

ROR Institution

Stanford University - https://ror.org/00f54p054

iD Principal Investigator

Mark A. Musen - https://orcid.org/0000-0003-3325-793X

Publications (1 .. ∞)

Modeling community standards for metadata as templates makes data FAIR. - https://pubmed.ncbi.nlm.nih.gov/36371407/

What have I kept hearing since that SAB meeting?

- Ontologies are “old fashioned”
- Ontology languages are too complicated
- Ontology engineering is not an interesting problem
- Ontologies coerce people to adopt rigid world views
- Generative AI obviates the need for ontologies

What can we do to quell these concerns?

- Acknowledge that ontology engineering is hard, but that good tools can make it easier
- Accept that OWL may not be the “one ring to rule them all”
- Educate people about the importance of standards for data management
- Accept that important terminological standards come in forms that are not ontological

And, despite a changing world, recognize that

- Symbolic representations will always play a major role in organizing data, information, and knowledge
- There is something intellectually important in the creation of symbolic models that help us to understand our world
- Precision in information management demands precise ontologies
- LLMs will never obviate the need for access to terminological standards

And, despite a changing world, recognize that

- Symbolic representations will always play a major role in organizing data, information, and knowledge
- There is something intellectually important in the creation of symbolic models that help us to understand our world
- Precision in information management demands precise ontologies
- LLMs will never obviate the need for access to terminological standards
- The Ontology Summit persists as an institution, and we have a lot to be proud of