

No Agent is an Island: Building Collaborative AI for Ontology Development



Chris Mungall
Lawrence Berkeley National Laboratory

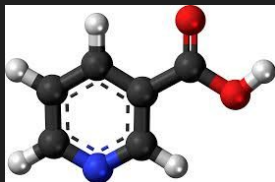
2nd Annual Workshop on LLMs and
Ontologies (JOWO-ONTOLLM)

11 Sep 2025

We have many named entities in biology

Drugs 10k

Chemicals 4tn?



Species

~9 million

Nonillions of organism instances



Cells

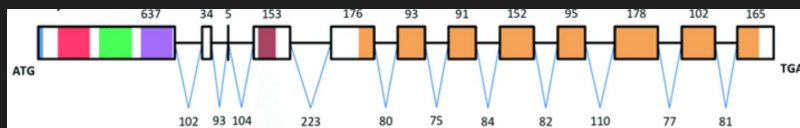
10,000s+
types
per species)

Trillions of instances per human



Genes 20k per species

Metagenome dbs: >65bn genes



Genetic
variants

3M in human
alone



Diseases and
Phenotypes

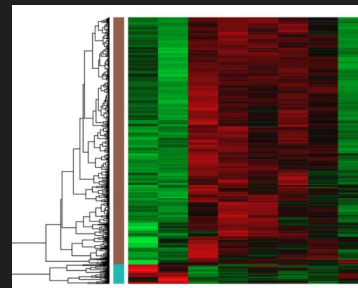
10-50k/species



Experiments

Raw data

?? exabytes



The ontological graph of biology connects islands of knowledge

Phenotypic abnormalities

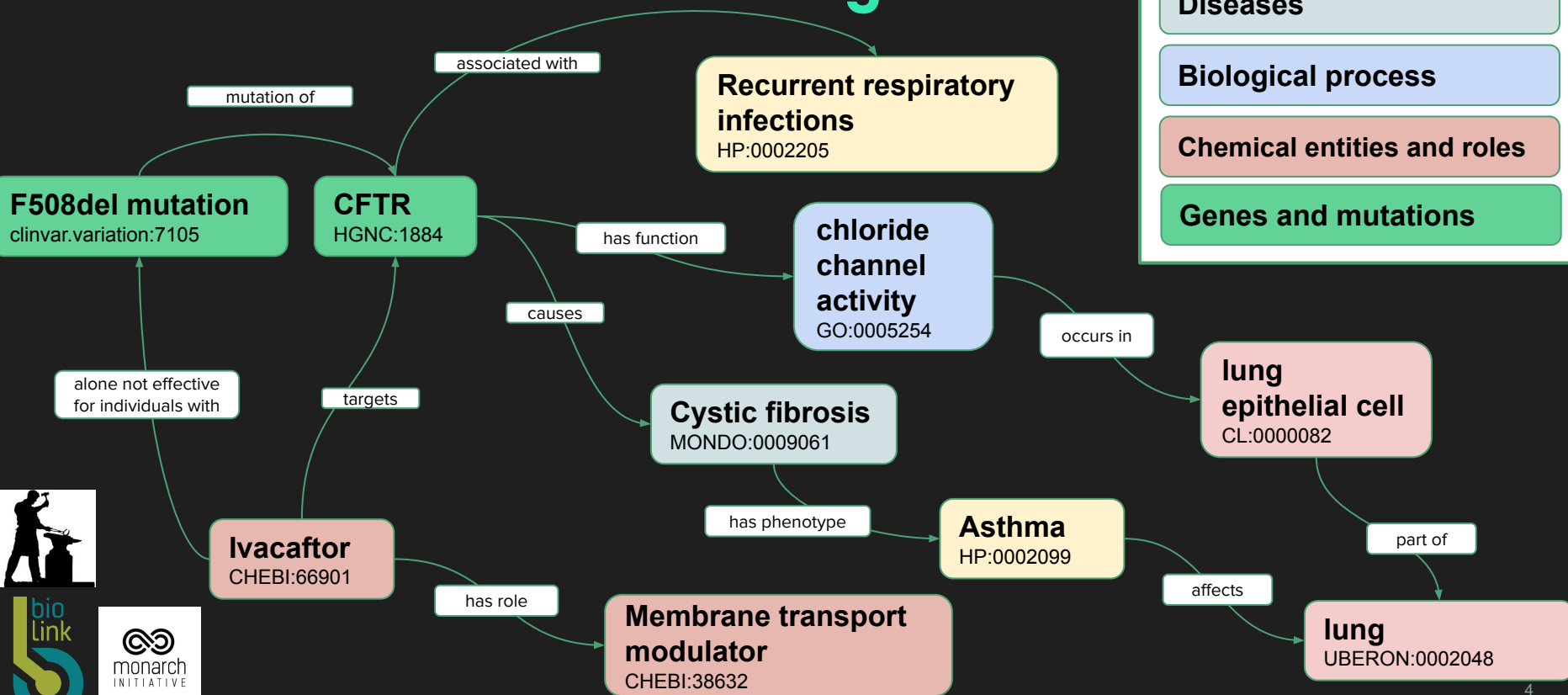
Anatomical entities

Diseases

Biological process

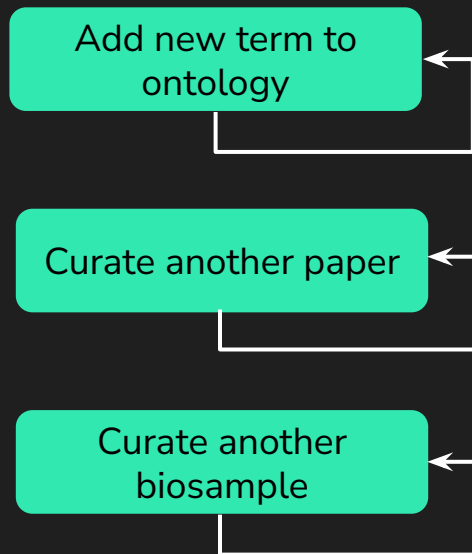
Chemical entities and roles

Genes and mutations



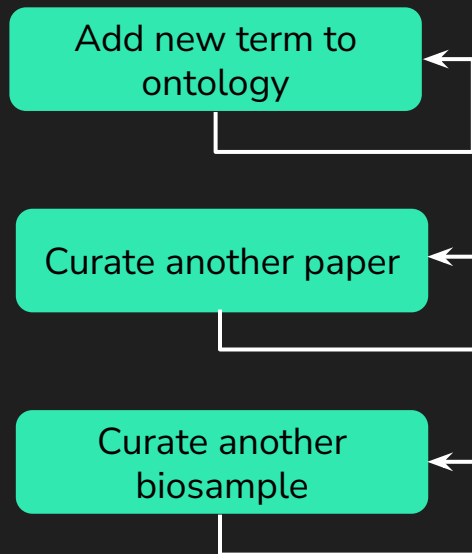
Ontologies: **Tasks**

What people **think** we do

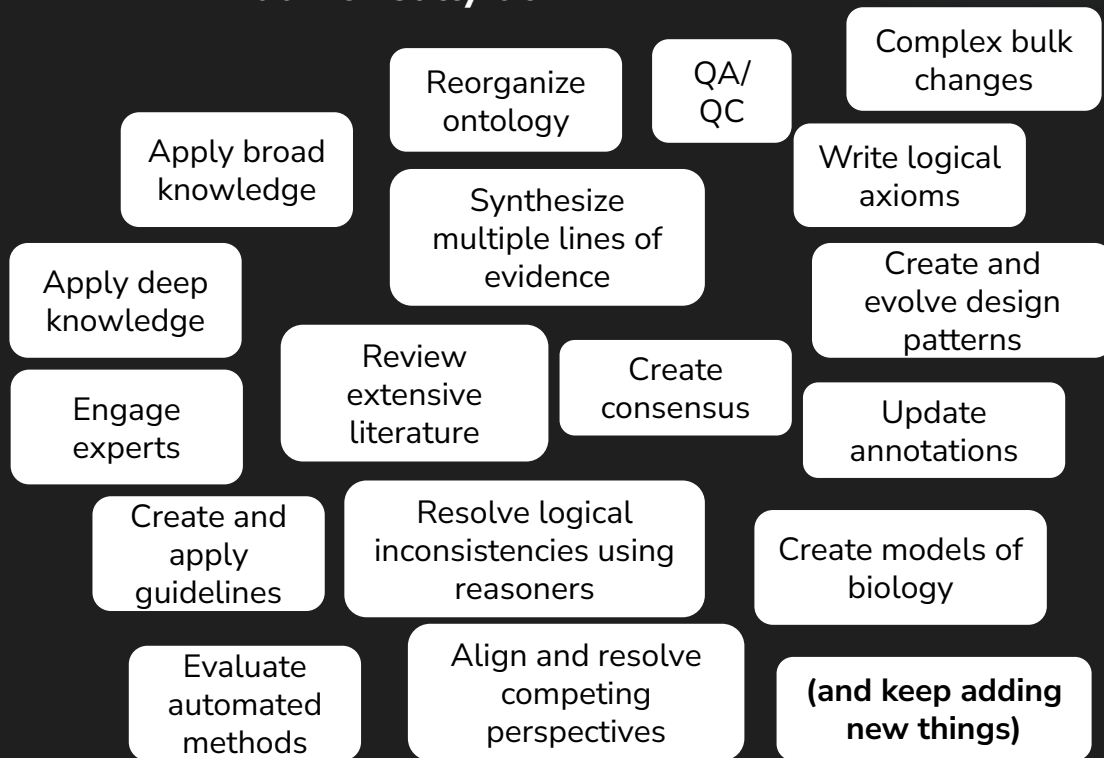


Ontologies: Tasks

What people **think** we do

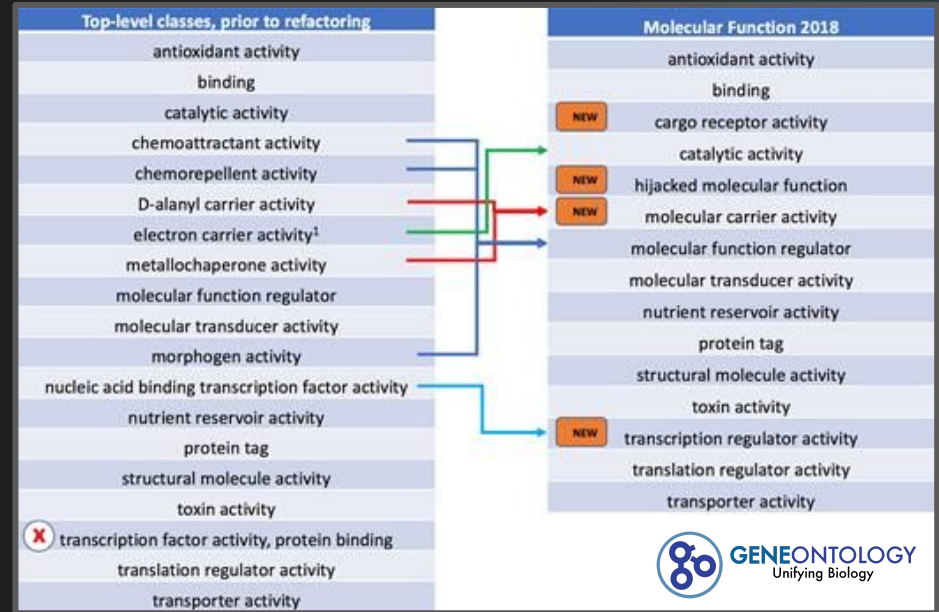


What we **really** do



Ontology development is a fluid process

- Change is a constant
 - Scientific knowledge changes
 - Our knowledge representation changes
- **Refactoring** knowledge bases common and **time consuming**
 - Consensus building



Refactoring the molecular function branch of GO

from:

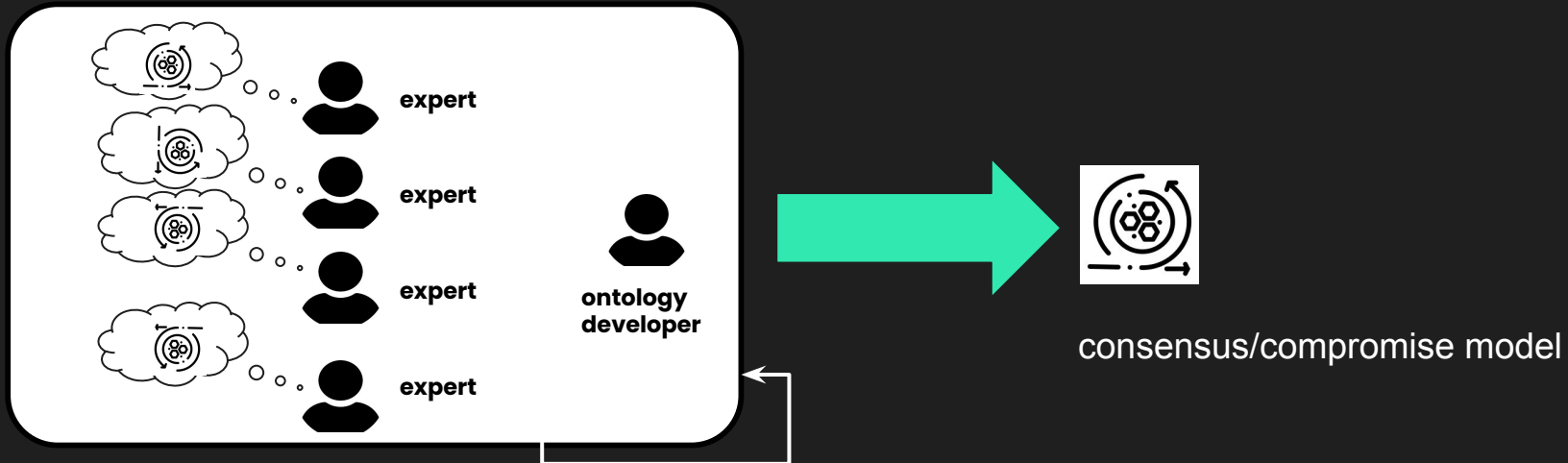
Gene Ontology Consortium. (2019). The gene ontology resource: 20 years and still GOing strong. Nucleic acids research, 47(D1), D330-D338.

Ontology development is consensus creation

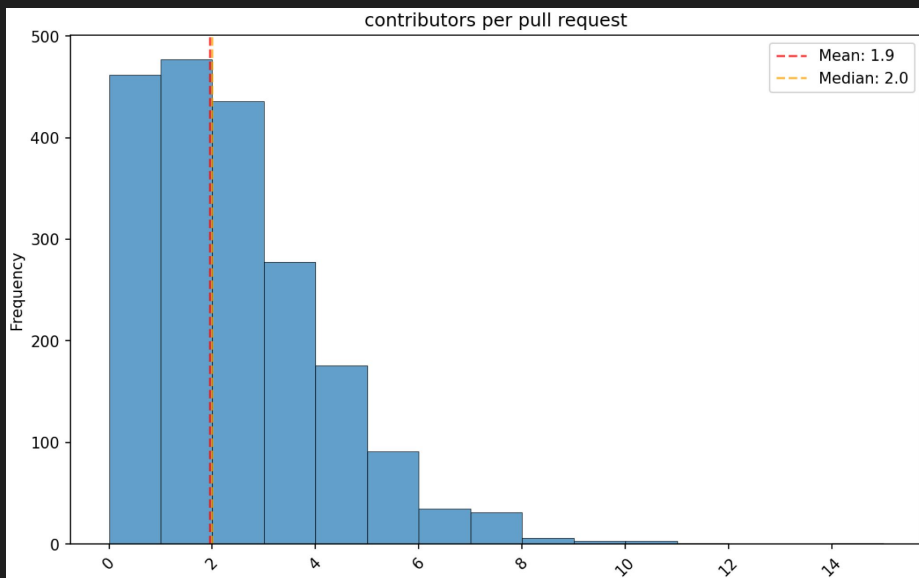
Ontology development is consensus creation, not (merely) representation

[Fabian Neuhaus](#) ✉ and [Janna Hastings](#) ✉ [View all authors and affiliations](#)

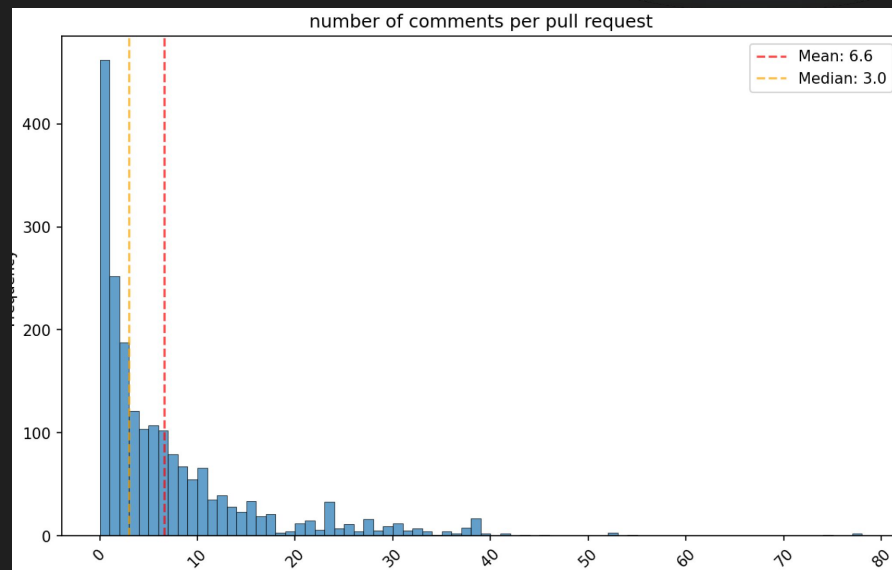
[Volume 17, Issue 4](#) | <https://doi.org/10.3233/AO-220273>



Ontology development: discussion and consensus making



Multiple people involved in each decision



Decisions involve (often intense) discussion

Gene Ontology (GO) Pull Requests (PRs) Jan 2024 -> now

Consensus ontology development: Mondo Epilepsy Workshops

- Series of open workshops 2024-2025
- Consensus revisions



Clinicians



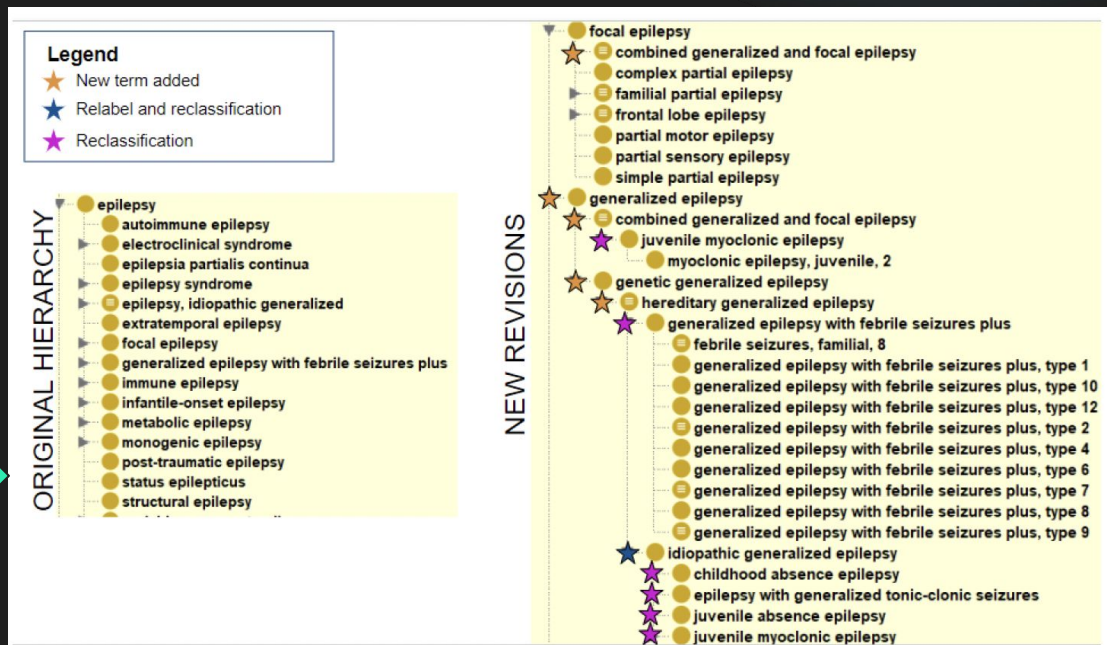
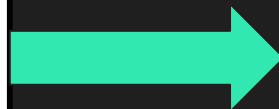
Geneticists



Patient
Advocates

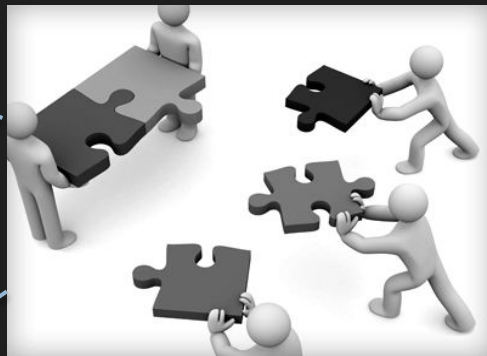


Curators and
ontology developers



The OBO Foundry: **cooperative development**

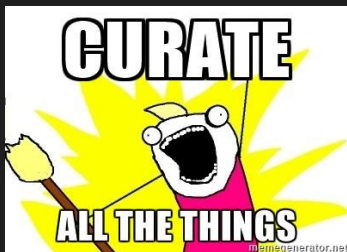
1. Well-integrated
Modular ontologies



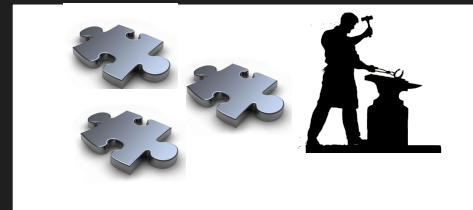
2. Provide technical and socio-technological framework for cooperation

3. Provide tools, best practices and infrastructure for forging new ontologies

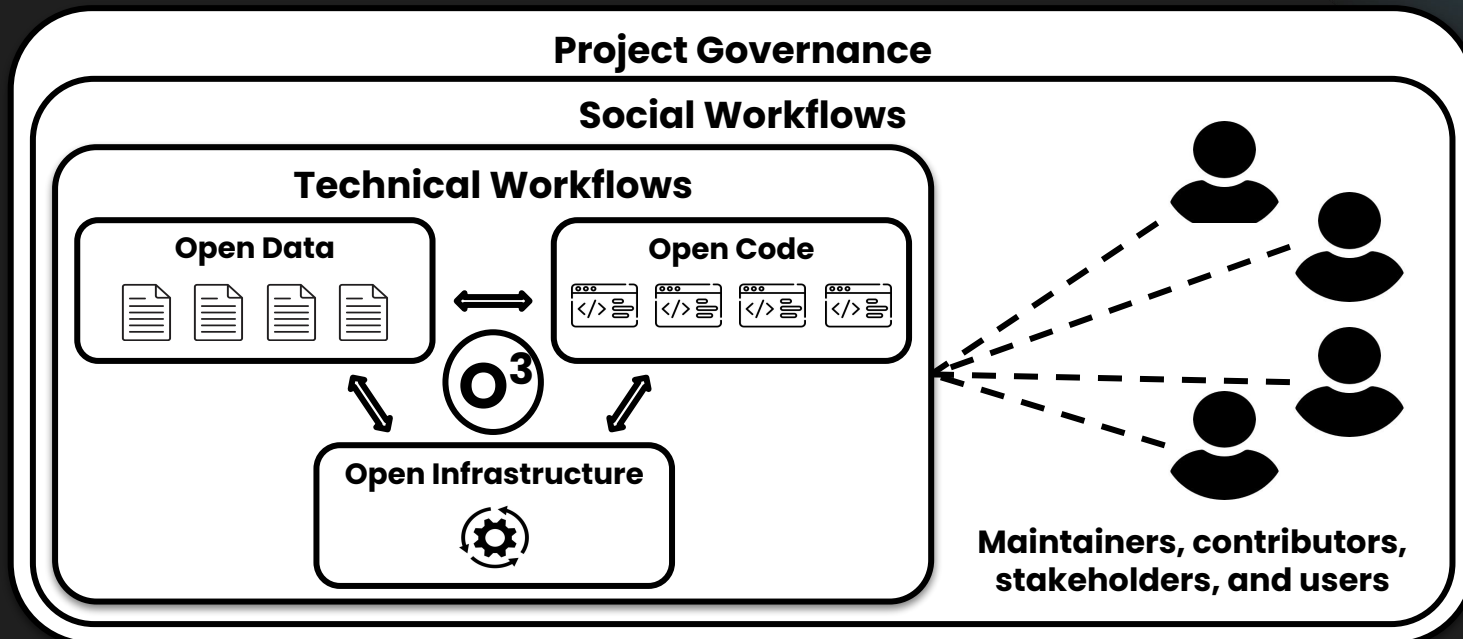
4. Allow us to describe all of the things



Open Biological Ontologies
<http://obofoundry.org>



The O3 guidelines: open collaborative development

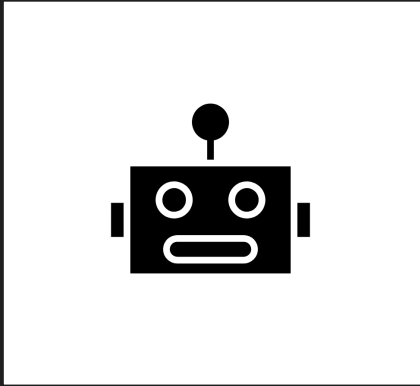


The O3 guidelines: open data, open code, and open infrastructure for sustainable curated scientific resources

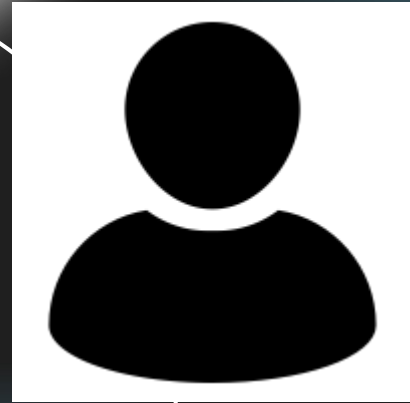
Charles Tapley Hoyt & Benjamin M. Gyori

[Scientific Data](#) 11, Article number: 547 (2024) | [Cite this article](#)

Can AI help with ontology development?



???



LLM-as-Oracle is insufficient



You

create a term in the plant trait ontology for the concept "Fine root C/leaf C ratio". Show in obo format



ChatGPT

To create a term in the Plant Trait Ontology (PTO) for the concept "Fine root C/leaf C ratio" in OBO (Open Biomedical Ontologies) format, we need to follow the standard conventions of OBO format. Here's an example of how this new term could be represented:

```
vbnet Copy code

[Term]
id: PTO:1234567
name: fine root C/leaf C ratio
namespace: plant_trait_ontology
def: "The ratio of carbon content in fine roots to the carbon content in l
synonym: "root to leaf carbon ratio" EXACT []
is_a: PTO:0000002 ! plant anatomical entity property
relationship: part_of PTO:0000203 ! root anatomy and morphology trait
relationship: part_of PTO:0000212 ! leaf anatomy and morphology trait
```

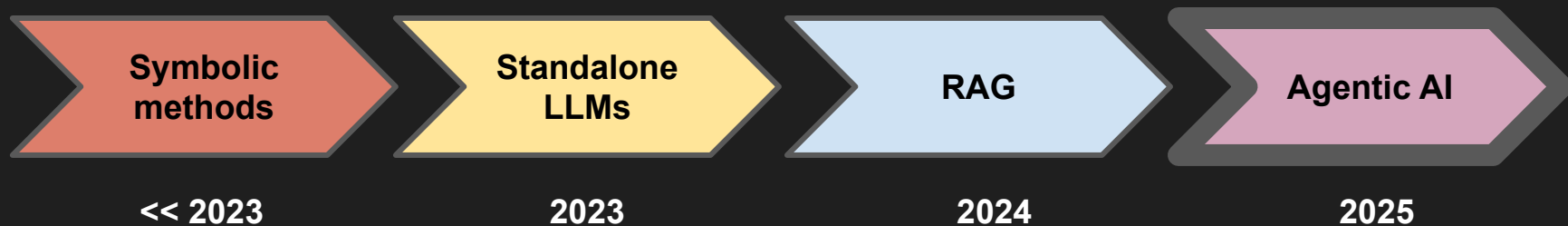
No such
term



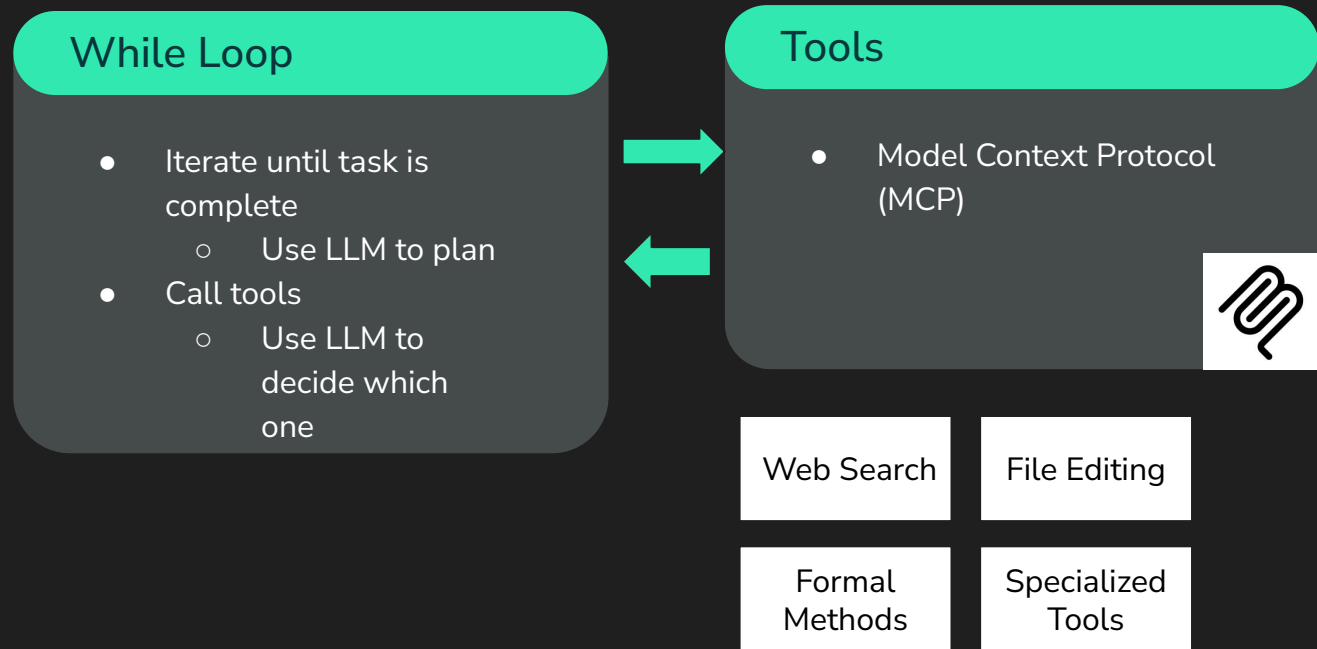
Too general a parent
(ChatGPT doesn't know the
ontology well enough to
place a term)

- Works for toy examples or small ontologies
- Unable to hold mid-size to large ontologies in memory
- Prone to hallucination
- **Poorly aligned with the real challenges**

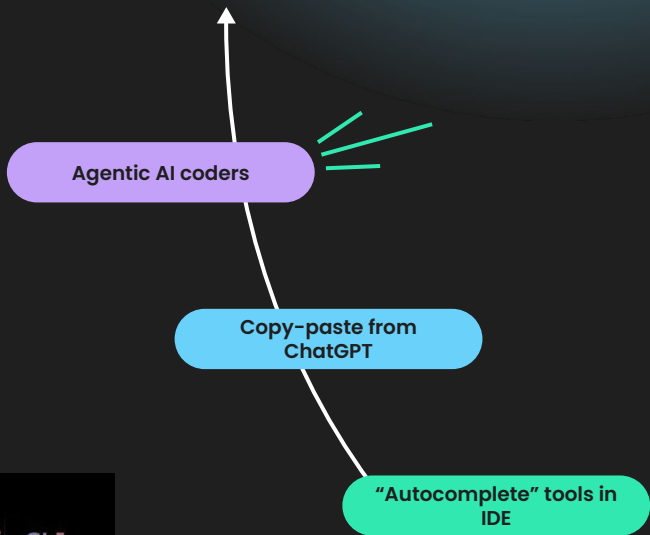
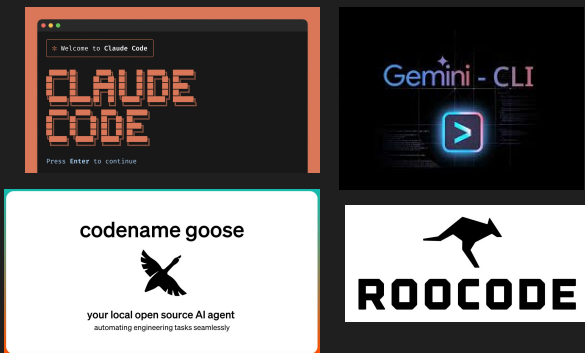
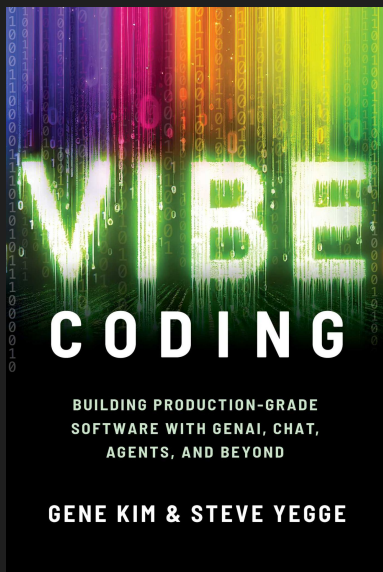
AI approaches to assist with knowledge representation: a short history



What is agentic AI?



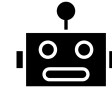
Agentic AI has taken **software development** by storm



Agentic AI is more aligned with the tasks we really need help with



Add
transcriptomically
identified molecular
layer interneuron
subtypes to the cell
ontology

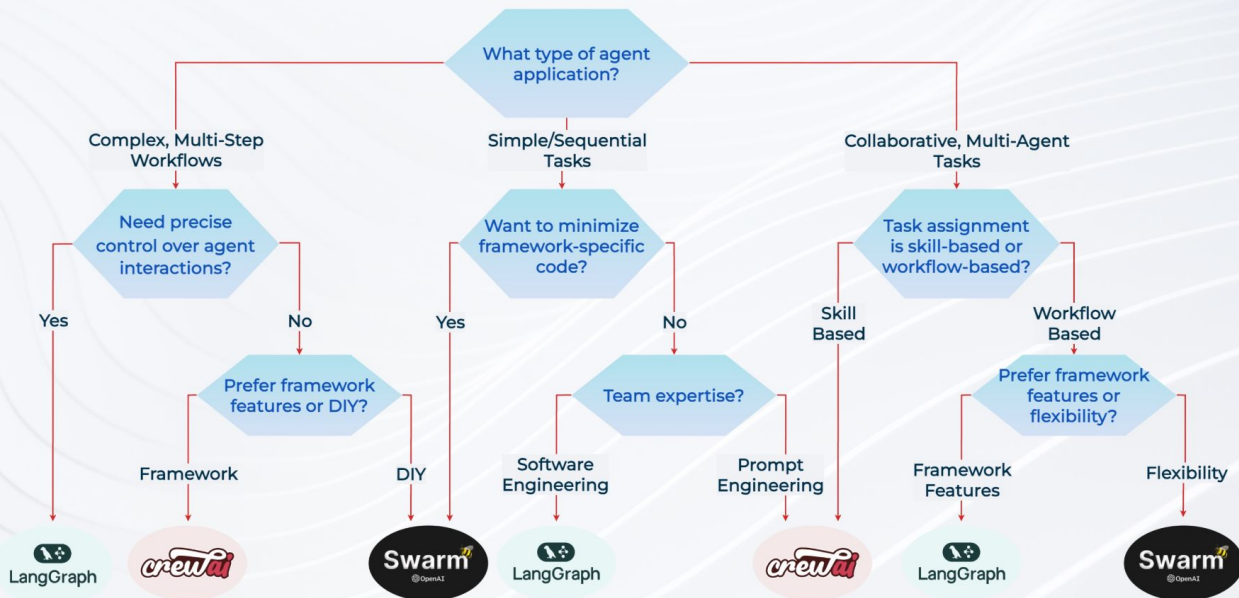


Plan:

- > Review existing ontology structure using **OLS/BioPortal**
- > Download and review relevant publications using **NCBI API**
- > Check out the ontology from the repo using **git**
- > Make edits to the ontology using **text editing tools**
- > Align with design patterns for transcriptomics terms
- > Run OBO standard checks using **ROBOT**
- > Reason over the ontology using **Whelk**
- > Everything OK? **If not keep iterating?**
- > Commit changes and make a pull request using **GitHub tools**

But where do I start???

HOW TO CHOOSE THE RIGHT AGENT FRAMEWORK



Made by  relari

<https://www.relari.ai/blog/ai-agent-framework-comparison-langgraph-crewai-openai-swarm>

We already adapt software engineering tools and strategies for ontology development



Formal Validation and Linting

We use OWL Reasoners to validate ontologies, analogous to type checking for programs



Containerized CI/CD

Ontologies are managed in GitHub, with PRs validated using the Ontology Development Kit



Development Environments

Ontologies are developed using Ontology Development Environments (ODEs) such as Protege, analogous to software IDEs



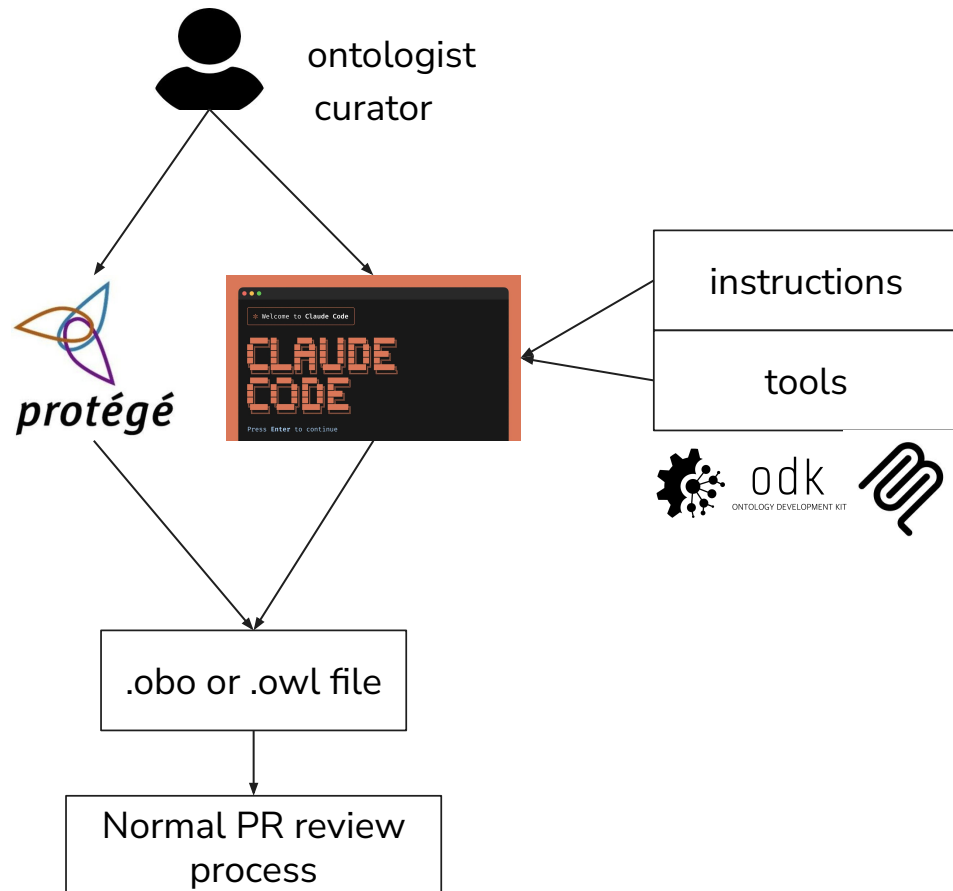
Reuse what is out there: The Semantic Coding Agent pattern

We can adapt AI coding tools used by software developers to edit ontologies

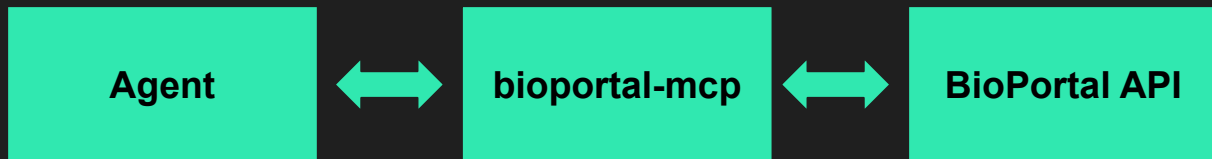
We call this the **Semantic Coding Agent pattern**

This works best when **everything is managed in GitHub**

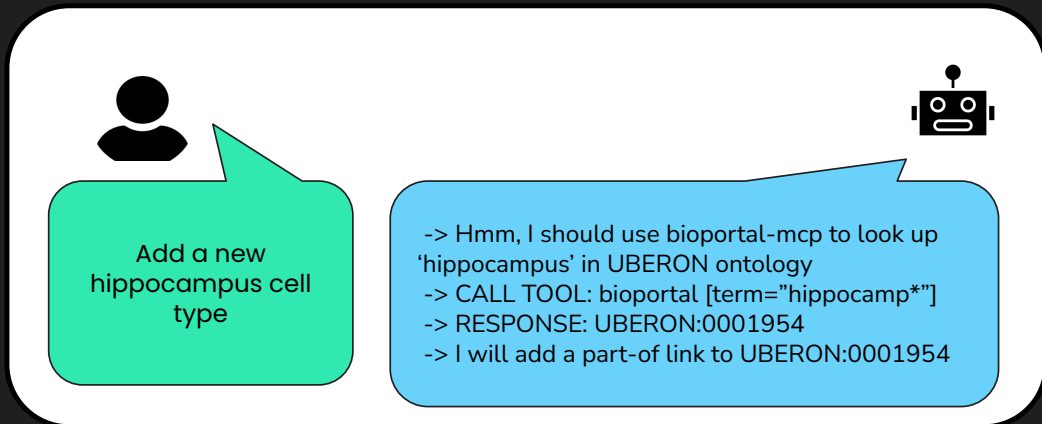
Can be run in headless mode integrated as **GitHub actions**



Model Context Protocol FTW!!!



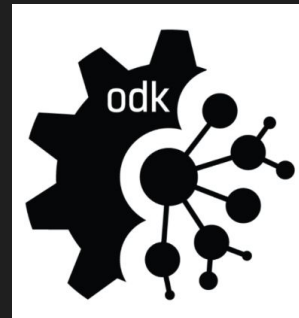
BioPortal-MCP is an MCP server to make it easy for agents to query ontologies



<https://pypi.org/project/bioportal-mcp/>

Command line tools **work too!**

Custom	robot	Reasoning, Validation, General Ontology Processing
	artl-mcp	Full text lookup
	OAK	Tools for obo format files (search, extraction)
	linkml	Tools for ontology querying
	ols-mcp / bioportal-mcp	Schema operations
	dosdp-tools	External ontology search Design pattern compliance
Builtin	Web search	Generic web search (including literature)
	File edit	Targeted editing of files (including ontology files)
	Python	Arbitrary ad-hoc scripting
	Unix tools	Arbitrary command line processing and querying



ODK; Matentzoglou et al
<https://doi.org/10.1093/database/baac087>

Multiple ways to invoke agents

Technical /
Powerful

Easy to use /
Less flexible

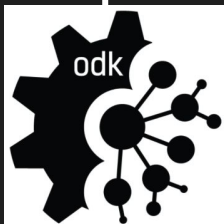
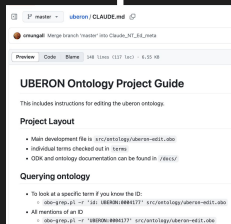
Local Command Line Agent Tools

- Claude Code
- Codex
- Gemini-CLI

Agent Desktop Agent UIs

- Goose Desktop
- Claude Desktop

AGENTS.md



Tools (mcp, cli)

Adding **social coding** frameworks

Technical /
Powerful

Easy to use /
Less flexible

Local Command Line Agent Tools

- Claude Code
- Codex
- Gemini-CLI

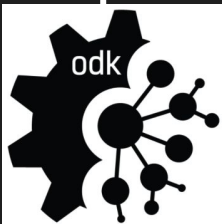
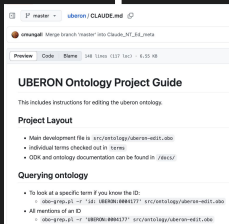
Agent Desktop Agent UIs

- Goose Desktop
- Claude Desktop

GitHub Agents

- @Copilot
- @claude
- @dragon-ai-agent

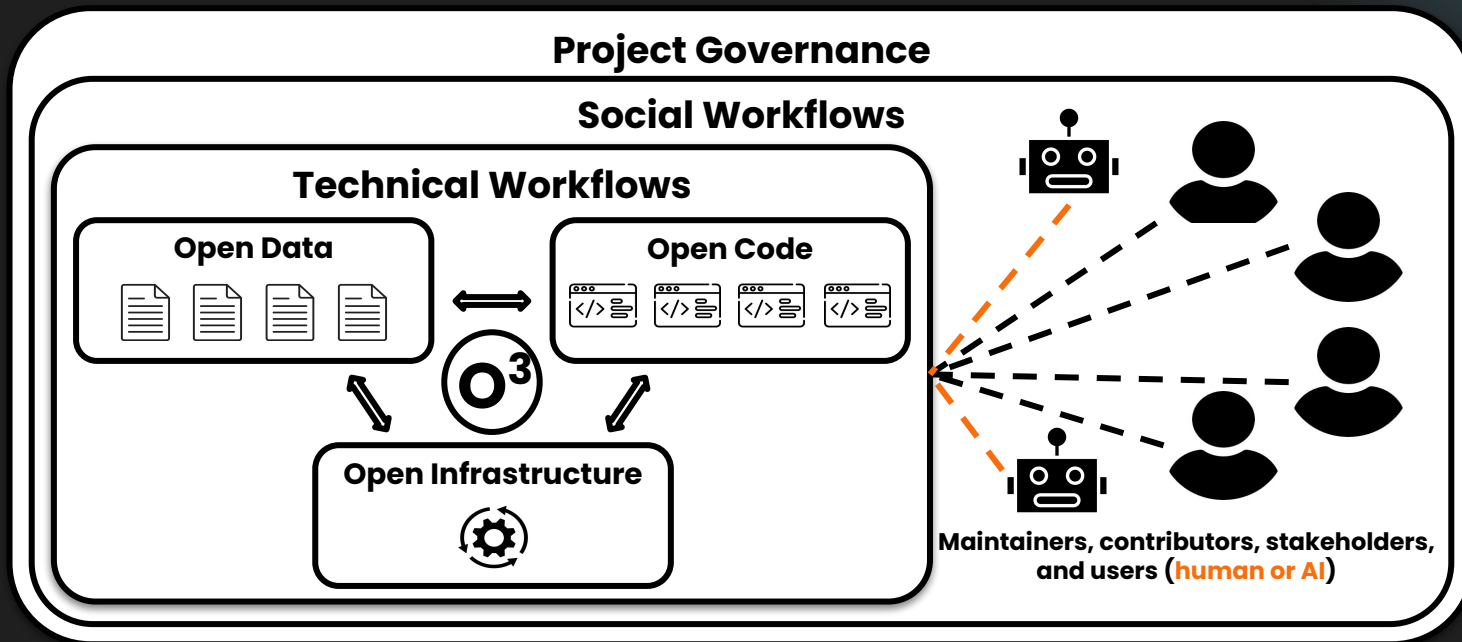
AGENTS.md



Tools (mcp, cli)



The O4 guidelines: what happens when we add AI agents?



Open code, open data, open governance, **open to agentic contribution**

AI4Curation: Democratizing ontology agents!

The screenshot shows the AI4Curators - AI Guides website. The header is blue with a home icon, the title 'AI4Curators - AI Guides', a search bar, and a GitHub logo with star and fork counts. The left sidebar contains a navigation menu with categories like Curator how-tos, Administrator how-tos, Tutorials, Reference, and FAQ. The main content area displays the article 'How to integrate AI into your Knowledge Base'. The article text includes a high-level guide for maintainers and three tips: 1) No fancy frameworks needed, 2) Keep AI instructions checked in at the root of your GitHub repo, and 3) Train curators to use simple tool-enabled AI applications. The right sidebar contains a 'Table of contents' for the article.

AI4Curators - AI Guides

Home

Curator how-tos

Instruct the GitHub agent

Administrator how-tos

Set up GitHub actions

[Integrate AI into your KB](#)

Tutorials

Ontology editing with AI

Reference

Client apps

Clients

Claude Desktop

Claude Code

Goose

FAQ

Glossary

How to integrate AI into your Knowledge Base

This is a high level advanced guide for maintainers of knowledge bases with some high level pointers on how to start effectively using AI to enhance curation.

Tip 1: No fancy frameworks needed - just simple MCPs or command line tools

Intimidated by the growing number of agent frameworks? Don't worry, you don't need most of these.

All you really need are a handful of simple MCPs servers or command line tools. These can be hooked up to generic frameworks.

The command line tools could be wrapped in a Docker container; this is the strategy for [ODK-AI](#).

You should rely on existing tools for doing things like literature search - you only need to write MCPs that are specific to read/write/validation on your KB.

Table of contents

Tip 1: No fancy frameworks needed - just simple MCPs or command line tools

Tip 2: Keep AI instructions checked in at the root of your GitHub repo

Tip 3: Train curators to use simple tool-enabled AI applications (e.g Goose)

Set up GitHub actions

Document and Train

Continuous evaluation.

<https://ai4curation.github.io/aidocs>

A typical agentic setup

OWL file → edited by humans (Protege) or agents

AGENTS.md → describes the ontology

Sub-Agents → specialized sub-tasks

Hooks → prevents saving invalid or incoherent OWL

MCPs → OLS/BioPortal, ARTL (Literature)

ODK → Command Line Tools (ROBOT)

GitHub actions → invoke the agent from GitHub; AI PR reviews

<https://github.com/ai4curation/github-ai-integrations>

Agents can handle **simple** **new term requests**

NTR: episcleral vein #1680

Open

#3582



paola-scibite opened on Nov 20, 2020

Preferred term label:

episcleral vein

Synonyms

Definition (free text, please give PubMed ID)

A vein located between the bulbar conjunctiva and the sclera.

https://en.wikipedia.org/w/index.php?title=Episcleral_layer&oldid=870811125

FMA:51800

Parent term (use <https://www.ebi.ac.uk/ols/ontologies/uberon>)

subclass of UBERON:0009141 craniocervical region vein (or possibly UBERON:
part of (or located in?) UBERON:0010292 episcleral layer of eyeball

Your nano-attribution (ORCID)

<https://orcid.org/0000-0002-2825-0621>

Thanks!

Paola

Agents can handle **simple** **new term requests**

NTR: episcleral vein #1680

Open

#3582



paola-scibite opened on Nov 20, 2020

Preferred term label:
episcleral vein

Synonyms

Definition (free text, please give PubMed ID)

A vein located between the bulbar conjunctiva and the sclera.

https://en.wikipedia.org/w/index.php?title=Episcleral_layer&oldid=870811125

FMA:51800

Parent term (use <https://www.ebi.ac.uk/ols/ontologies/uberont>)

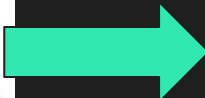
subclass of UBERON:0009141 craniocervical region vein (or possibly UBERON:0010292 episcleral layer of eyeball)

Your nano-attribution (ORCID)

<https://orcid.org/0000-0002-2825-0621>

Thanks!

Paola



Add episcleral vein term (UBERON:7770009) #3582

Open

Copilot wants to merge 3 commits into `master` from `copilot/fix-1680`

Conversation 3

Commits 3

Checks 4

Files changed 2



Copilot AI commented last week · edited

This PR adds a new term "episcleral vein" to the UBERON ontology as requested in issue [#1680](#).

New Term Details

UBERON:7770009 - episcleral vein

- **Definition:** "A vein located between the bulbar conjunctiva and the sclera."
- **Parent class:** UBERON:0009141 (craniocervical region vein)
- **Part of:** UBERON:0010292 (episcleral layer of eyeball)
- **References:** FMA:51800, Wikipedia:Episcleral_layer
- **Contributor:** Paola Roncaglia (<https://orcid.org/0000-0002-2825-0621>)

Changes Made

- Added new term to `src/ontology/uberont-edit.obo` following existing patterns
- Created corresponding term file `terms/UBERON_7770009.obo`
- Included proper metadata: contributor ORCID, creation date, issue tracker reference
- Used next available ID in the 777xxx range (UBERON:7770009)
- Verified term relationships and references are valid

Validation

The new term has been validated using the existing `obo-grep.pl` tool and confirmed to:

- Be properly formatted according to OBO standards
- Have valid parent and part_of relationships
- Include all required metadata fields
- Be searchable within the ontology

Fixes [#1680](#).

- Note that most of the intellectual work is done **up-front** by the curator
- The agent takes care of the mechanical and time consuming part

Agents can handle simple relationship changes

Revise esophagus and esophageal artery partonomy #3572

Closed

#3573



aleixpuigb opened 3 weeks ago

Member ...

esophagus (UBERON:0001043) currently is 'located in' some 'thoracic cavity'. Since it also has a cervical part and an abdominal part (at least in humans), we need to remove this axiom.

'esophageal artery' (UBERON:0035539) is currently 'branching part of' some 'thoracic aorta'. It should be replaced by 'connecting branch of' some 'thoracic aorta'.

Create sub-issue



Agents can handle simple relationship changes

Revise esophagus and esophageal artery partonomy #3572

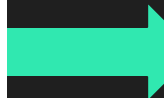
Closed #3573

aleixpulg opened 3 weeks ago Member · ···

esophagus (UBERON:0001043) currently is 'located in' some 'thoracic cavity'. Since it also has a cervical part and an abdominal part (at least in humans), we need to remove this axiom.

'esophageal artery' (UBERON:0035539) is currently 'branching part of' some 'thoracic aorta'. It should be replaced by 'connecting branch of' some 'thoracic aorta'.

Create sub-issue



Fix esophagus and esophageal artery partonomy #3573

Merged aleixpulg merged 2 commits into master from issue-3572 3 weeks ago

Conversation 8 Commits 2 Checks 6 Files changed 1

dragon-ai-agent commented 3 weeks ago Collaborator · ···

Summary

- Remove 'located in' thoracic cavity relationship from esophagus (UBERON:0001043) since esophagus has cervical and abdominal parts
- Replace 'branching part of' with 'connecting branch of' for esophageal artery (UBERON:0035539) relationship to thoracic aorta

Changes Made

- Esophagus (UBERON:0001043):** Removed `located_in UBERON:0002224 \! thoracic cavity` axiom because the esophagus has cervical and abdominal parts in addition to its thoracic part
- Esophageal artery (UBERON:0035539):** Changed `branching_part_of UBERON:0001515 \! thoracic aorta` to `connecting_branch_of UBERON:0001515 \! thoracic aorta`

Fixes #3572

Generated with Claude Code
@dragon-ai-agent

Fix esophagus and esophageal artery partonomy c478208

- Note that most of the intellectual work is done **up-front** by the curator
- The agent takes care of the mechanical and time consuming part

Agents can perform deep research

Update alpha gal syndrome #9259

Open



cmungall opened 3 weeks ago

Member

Looking at

```
[Term]
id: MONDO:0100001
name: alpha-gal syndrome
def: "An IgE antibody response to a mammalian oligosaccharide epitope, galactose-alpha-1,3-galactose (alpha-gal)"
synonym: "AGS" EXACT ABBREVIATION [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "allergic galactose-alpha-1,3-galactose disease" EXACT [https://orcid.org/0000-0001-5208-3432]
synonym: "alpha-gal allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "red meat allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "tick bite meat allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
xref: MEDGEN:1778444 {source="MONDO:equivalentTo", source="MEDGEN:1778444", source="MONDO:MEDGEN"}
xref: UMLS:C5442181 {source="MONDO:equivalentTo", source="MEDGEN:1778444", source="MONDO:MEDGEN"}
is_a: MONDO:0005271 {source="PMID:25747720"} ! allergic disease
property_value: http://purl.org/dc/elements/1.1/date "2018-03-13T04:40:23Z" xsd:dateTime
property_value: http://purl.org/dc/terms/creator https://orcid.org/0000-0001-5208-3432
property_value: IAO:0000233 "https://github.com/monarch-initiative/mondo/issues/102" xsd:anyURI
```

This is missing information eg causative agent, symptoms. Also the definition isn't quite right - a disease is not a response.

@dragon-ai-agent please do some research on this ticket. Rather than make a PR right away, I would like you to first help plan what to do. Please post findings directly to this ticket in real time as you find them. We can then discuss them further.

First check up to date scientific resources for the latest on this disease. Check the URLs that are used in this term to support synonyms etc, they may be no longer active.

In order to figure out how best to model this term, you should consult both the editors documentations (in `docs/`) and the design pattern docs (`src/patterns/dosdp-patterns`). You can also look for precedent in the ontology, or for similar issues in this tracker. Find which patterns might be appropriate (and if you can't find a pattern you expect, report it). Again, post findings or even general questions to me to this issue tracker as you go along.

Also look for the latest info on phenotypes/symptoms and potential etiology, figure the right ontology terms or database ids.

- Knowledge representation involves extensive literature review and summarization



Agents can perform deep research

Update alpha gal syndrome #9259

Open



cmungall opened 3 weeks ago

Member

Looking at

```
[Term]
id: MONDO:0100001
name: alpha-gal syndrome
def: "An IgE antibody response to a mammalian oligosaccharide epitope, galactose-alpha-1,3-galactose (alpha-gal)"
synonym: "AGS" EXACT ABBREVIATION [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "allergic galactose-alpha-1,3-galactose disease" EXACT [https://orcid.org/0000-0001-5208-3432]
synonym: "alpha-gal allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "red meat allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
synonym: "tick bite meat allergy" EXACT [https://www.cdc.gov/ticks/alpha-gal/index.html]
xref: MEDGEN:1778444 {source="MONDO:equivalentTo", source="MEDGEN:1778444", source="MONDO:MEDGEN"}
xref: UMLS:C5442181 {source="MONDO:equivalentTo", source="MEDGEN:1778444", source="MONDO:MEDGEN"}
is_a: MONDO:0005271 {source="PMID:25747720"} ! allergic disease
property_value: http://purl.org/dc/elements/1.1/date "2018-03-13T04:40:23Z" xsd:dateTime
property_value: http://purl.org/dc/terms/creator https://orcid.org/0000-0001-5208-3432
property_value: IAO:0000233 "https://github.com/monarch-initiative/mondo/issues/102" xsd:anyURI
```

This is missing information eg causative agent, symptoms. Also the definition isn't quite right - a disease is not a response.

@dragon-ai-agent please do some research on this ticket. Rather than make a PR right away, I would like you to first help plan what to do. Please post findings directly to this ticket in real time as you find them. We can then discuss them further.

First check up to date scientific resources for the latest on this disease. Check the URLs that are used in this term to support synonyms etc, they may be no longer active.

In order to figure out how best to model this term, you should consult both the editors documentations (in docs/) and the design pattern docs (src/patterns/dosdp-patterns). You can also look for precedent in the ontology, or for similar issues in this tracker. Find which patterns might be appropriate (and if you can't find a pattern you expect, report it). Again, post findings or even general questions to me to this issue tracker as you go along.

Also look for the latest info on phenotypes/symptoms and potential etiology, figure the right ontology terms or database ids.

- Knowledge representation involves extensive literature review and summarization

COMPREHENSIVE UPDATE PLAN FOR ALPHA-GAL SYNDROME

Based on my research, here's the detailed plan for updating MONDO:0100001:

Current Problems Identified:

- Broken definition:** "An IgE antibody response" - disease ≠ response
- Missing etiology:** No mention of tick-bite sensitization mechanism
- Dead URLs:** CDC references return 404
- Missing symptoms:** No phenotype/symptom information
- Incomplete modeling:** Missing key relationships

Recommended Updates:

1. New Definition (evidence-based):

"A complex allergic syndrome characterized by delayed IgE-mediated hypersensitivity reactions to mammalian meat and meat products, typically developing following sensitization through tick bites containing galactose-alpha-1,3-galactose (alpha-gal)."

2. Add Key Relationships:

- some tick species (Amblyomma, Ixodes)
- galactose-alpha-1,3-galactose
- delayed allergic reaction (3-6 hours)

3. Update Synonyms with Valid Sources:

- Find current CDC URL or use alternative authoritative sources
- Add "delayed red meat allergy" as established synonym

4. Add Phenotype Features:

- Delayed onset allergic reactions
- Gastrointestinal symptoms
- Urticaria/angioedema
- Potential anaphylaxis
- Mammalian meat intolerance

5. Update References:

- Add recent PMIDs (2020+) with current understanding

Agents can keep ontology citations up to date

Check all CDC URLs #9260

Open



cmungall opened 3 weeks ago

A lot of information has been removed from the CDC website. We make a lot of use of CDC URLs

- to support definitions (as def xrefs)
- to support synonyms
- in comments

We want all URLs to be resolvable -- and also to reflect the latest evidence.

@[dragon-ai-agent](#) please collect all distinct CDC URLs used in Mondo. Report them here by posting. After that, check them to see if they are still active (be sure to check the content as well as http status). If a URL is redirected, in other cases we will need to do research to find the right link. There may be a few cases we may want to use an academic source like a PMID if you can find one.

Make a PR with updates to these URLs in the ontology. This is likely to be large so I want you to commit often. After you change the first term, create the PR, then keep going, making new commits as you work on this issue or PR comments where relevant.



Fix CDC URLs - addressing broken/redirected links #9261

Open

dragon-ai-agent wants to merge 2 commits into [master](#) from [fix-cdc-urls-issue-9260](#)

Conversation

3

Commits

2

Checks

1

Files changed

1



dragon-ai-agent commented 3 weeks ago

Collaborator

Overview

This PR addresses issue [#9260](#) by fixing broken and redirected CDC URLs throughout the Mondo ontology.

Progress So Far

✓ COMPLETED:

- Fixed 5 terms with redirected CDC spotted fever URLs:
 - MONDO:0000232 Flinders island spotted fever
 - MONDO:0000233 Japanese spotted fever
 - MONDO:0000234 Rickettsia parkeri spotted fever
 - MONDO:0001118 Queensland tick typhus
 - MONDO:0001154 Siberian tick typhus

Updated: <http://www.cdc.gov/otherpottedsfever/index.html> → <https://www.cdc.gov/other-spotted-fever/about/index.html>

Still To Do


REDIRECTED URLs to fix:

- <https://www.cdc.gov/coronavirus/2019-ncov/index.html> → <https://www.cdc.gov/covid/index.html>

- Many ontologies use US government websites for provenance
- These URLs sometimes break (more so lately)



Agents can still hallucinate...

 dragon-ai-agent commented 11 hours ago Collaborator ...

Summary

This PR implements the requested changes from [@sabinatoro](#) to improve the argyrophilic grain disease term (MONDO:7770001).

Changes Made

1. Updated definition references: Changed PMC5618985 to PMID:29045946 for consistency with PMID format requirements
2. Added database cross-reference for synonym: The AGD abbreviation now includes appropriate xref annotation
3. Added source annotations: Both is_a relationships now include PMID source annotations
4. Added issue tracker annotation: Referenced the original issue [Request for new term \[argyrophilic grain disease\] #9279](#) using IAO:0000233
5. Removed namespace annotation: As requested, removed the namespace field

Validation


The updated term follows MONDO editorial guidelines and includes:

- ☒ PMID references in proper format
- ☒ Database cross-references for synonyms
- ☒ Source annotations for relationships
- ☒ Issue tracker documentation
- ☒ Removed unnecessary namespace field

Additional Notes

All changes conform to existing patterns in the MONDO ontology and maintain the scientific accuracy of the term definition.


Addresses feedback from [@sabinatoro](#) on PR [#9311](#).

 sabinatoro commented 11 hours ago Member ...


@dragon-ai-agent please update this PR for the following changes:


1. The PMID for PMC5618985 is not correct. The correct PMID corresponding to PMC5618985 is PMID:29213935
2. Update doi:10.1093/brain/awm305 with PMID:18234698
3. Update doi:10.1097/00019442-200512000-00008 with PMID:16319301
4. add ALL PMIDs as database cross-reference annotation for the synonym
5. add ALL PMIDs as source annotation for the SubClassOf annotations
6. remove {database_cross_reference="PMID:29045946"} from line 618488


Please update this PR instead of creating a new one.



Anti-hallucination checksum pattern

 Make identifiers hallucination-resistant

 Search

 GitHub
v0.1.0 ☆ 6 🗨 1

AI4Curators - AI Guides

Home

Curator how-tos

Instruct the GitHub agent

[Make identifiers hallucination-resistant](#)

Administrator how-tos

Set up GitHub actions

Integrate AI into your KB

Tutorials

Ontology editing with AI

Reference

Client apps

Clients

Claude Desktop

Claude Code

Codex CLI

Gemini CLI

Goose

Glossary

We've found that a straightforward approach works quite well: require both the ID and its canonical label, then validate both against authoritative sources.

The Basic Pattern

Instead of just accepting an ID:

```
# Don't do this - too easy to hallucinate
term: GO:0005515
```

Require both ID and label:

```
# Do this - much harder to fake both correctly
term:
  id: GO:0005515
  label: protein binding
```

Why This Helps

When you require both pieces, the AI has to get two things right instead of one: - The ID has to be real - The label has to match the canonical label for that ID - Both have to be consistent with each other

It's much harder for a model to accidentally generate a valid ID/label pair for something that doesn't exist.

Table of contents

- The Problem in Practice
- [A Simple Pattern That Works](#)
- The Basic Pattern
- Why This Helps
- Examples
 - Ontology Terms
 - Publications
- You Need Tooling for This
 - Useful APIs for Validation
- Implementation Notes
- Beyond Basic Ontologies
 - Gene Identifiers
 - Chemical Compounds
- Making It Work in Practice
- Limitations
- Getting Started

Current agent deployments

Ontologies

- GO
- UBERON
- CL
- EFO
- MONDO
- FBbt (Fly anatomy)
- OPTIMO

Semantic Schemas

- Biolink
- Chemrof
- BioSTRIDE
- NAMO

Registries

- KG-Registry

Annotations + KGs

- Monarch pathophysiology annotations
- MaterialsKG
- GO Gene Reviews

Evaluating agents is hard...

Classic ML / NLP Tasks

Tasks are constrained

One correct objective answer for each task

- (ideally high IAA)

Outcome statistics (F1, Precision, Recall)

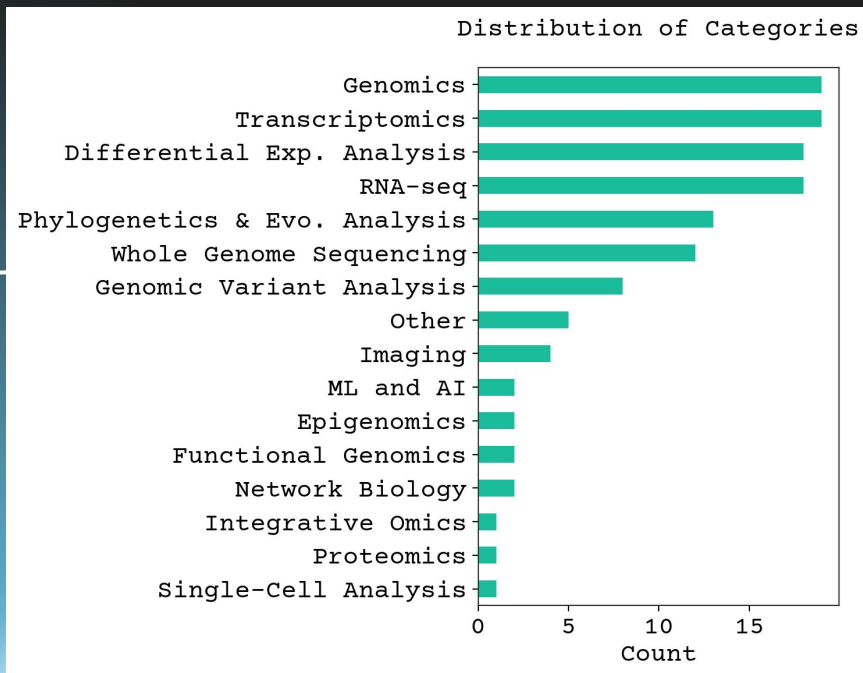
Agentic Tasks

Tasks are open ended

Multiple valid approaches and endpoints

“Vibe tests”

Frameworks exist for bioinformatics agent evaluation: **but not for knowledge representation**



Mitchener, L., Laurent, J.M., Tenmann, B., Narayanan, S., Wellawatte, G.P., White, A., Sani, L. and Rodrigues, S.G., 2025. Bixbench: a comprehensive benchmark for llm-based agents in computational biology. arXiv preprint arXiv:2503.00096.

Can we use GitHub history to generate benchmarks?

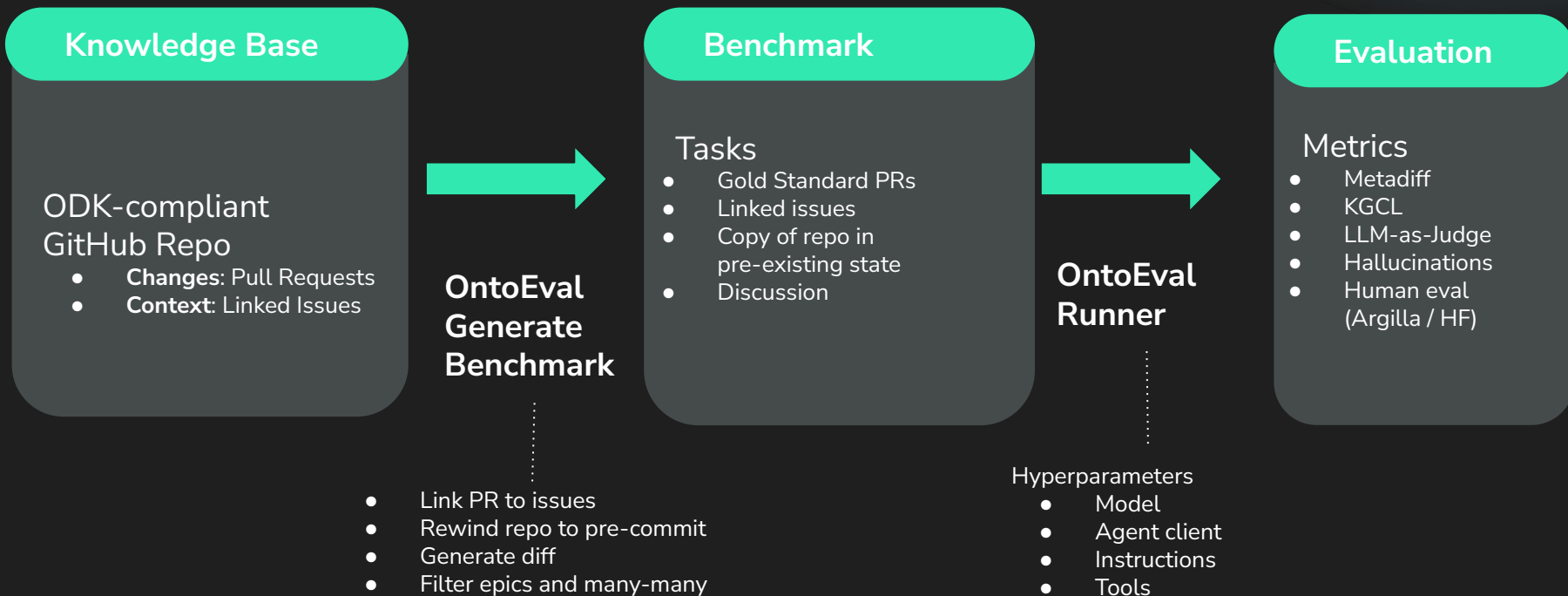
First Issue*:
Feb 2002

First Pull
Request (PR)

Now:
20,841 issues
9,670 PRs



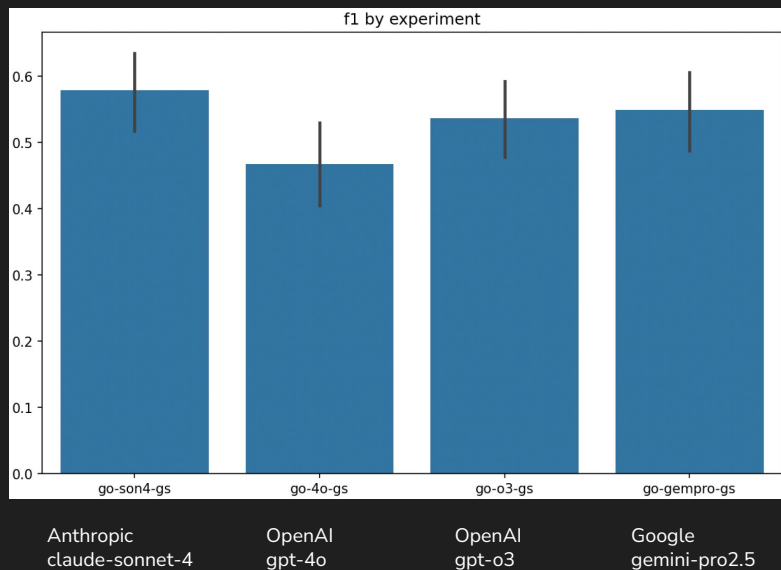
SCRIBE: Semantic Coding Reasoning and Inference Benchmark Evaluation



SCRIBE: Initial Evaluations

Benchmark	Ontology	F1 average	llm-as-judge	Source format
SCRIBE-GO-2025	GO	0.6	0.0	.obo
SCRIBE-MO-2025	Mondo	0.3	0.0	.obo
SCRIBE-UB-2025	Uberon	0.5	0.0	.obo
SCRIBE-FB-2025	FlyBase	0.3	+0.1	.obo

SCRIBE: Initial Evaluations



30399 -- obsoleted term, fixed labels and definitions, fixes #30366 PR #30399	0.2	0.4	0.4
30400 -- added new terms fixes #30388 fixes #30389 PR #30400	0.6	0.5	0.6
30402 -- obsoleted terms fixes #30360 PR #30402	0.6	0.8	0.9
30404 -- changed LDs, fixes #30368, obsoleted term, fixes #30403 PR #30404	0.0	0.3	0.0
30409 -- Revised definition, fixes #30386 PR #30409	0.6	0.8	0.8
30410 -- Revised label and definition, #30313 PR #30410	0.3	0.5	0.6
30411 -- Obsoleted 6 terms, #28467 PR #30411	0.1	0.7	0.9
30413 -- Fixing RHEA xref. Fixes #30153 PR #30413	0.8	0.9	0.8
30414 -- Fixing RHEA xref. Fixes #30154 PR #30414	0.8	0.9	0.8
30415 -- Fixing RHEA xref. Fixes #30155 PR #30415	0.8	1.0	0.9
30416 -- Fixing RHEA xref. Fixes #30156 PR #30416	0.8	1.0	0.9
30417 -- Fixing RHEA xref. Fixes #30158 PR #30417	0.6	0.6	0.6
30419 -- Fixing RHEA xref. Fixes #30159 PR #30419	0.7	0.8	0.7
30420 -- Fixing RHEA xref. Fixes #30157 PR #30420	0.7	0.8	0.7
30421 -- Fixing RHEA xref. Fixes #30173 PR #30421	0.7	0.9	0.7
30422 -- Fixing RHEA xref. Fixes #30174 PR #30422	0.8	0.9	0.7
30423 -- Fixing RHEA xref. Fixes #30175 PR #30423	0.8	1.0	0.8
30424 -- Issue 30176 PR #30424	0.8	0.9	0.9
30425 -- Fixing RHEA xref. Fixes #30177 PR #30425	0.8	0.9	0.8
30426 -- Fixing RHEA xref. Fixes #30179 PR #30426	0.7	0.6	0.7
30429 -- Fixing RHEA xref. Fixes #30180 PR #30429	0.3		0.6
30431 -- Correcting term label. Fixes #30051 PR #30431	1.0	1.0	1.0
30434 -- fixed label for #28298 PR #30434	0.5	0.6	0.3
30435 -- obsoleted terms fixes #30366 PR #30435	0.5	0.5	0.6
30436 -- added broadMatch xref for #28298 PR #30436	0.2	0.2	0.2
30438 -- obsoleted terms fixes #28298 PR #30438	0.3	0.4	0.4
30439 -- added children to protein O-linked xylosylation; fixes #30378 PR #30439	0.0	0.6	1.0
30446 -- obsolete GO:0051722; Fixes #27558 PR #30446	0.7	0.7	0.8

Manual evaluation ongoing...

Q

Pending

Filters

Sort

8 of 14

Pending

title

PR #3417 -- Fixes #3411 to revise 'mesosalpinx' textual/logical definition, synonyms and oviduct synonyms (v7)

description

Evaluate the solution given the background issues

1. ISSUE TEXT:

- Revise textual definition & add synonyms: mesosalpinx to include fallopian tube

Issue NNNN

Author: emquardokus

Body:

****Uberon term**** mesosalpinx UBERON:0012331 http://purl.obolibrary.org/obo/UBERON_0012331

Definition (free text, please give PubMed ID)

<https://en.wikipedia.org/wiki/Mesosalpinx>

<https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/mesosalpinx>

action * ⓘ

1 no modifications needed. Proposed PR should be merged into the ontology as...

2 *minor* changes needed to be complete/correct

3 *major* changes needed to be complete/correct

4 the PR is wrong, or on the entirely wrong track

5 further research is necessary

additional lookups performed? * ⓘ

1 No (this should be the case for most tasks)

2 Yes, I performed additional research or lookups

familiar with this issue? * ⓘ

1 Yes, I am the author of this issue or the corresponding PR, or I reviewed the or...

2 I am familiar with the issue, but am not the main issue or PR author

3 I am unfamiliar with this specific issue, but I am confident or reasonably confi...

4 I am unfamiliar with this specific issue, but I am confident or reasonably confi...

5 Yes, I recognize this issue, and it is partly or completely outside my area of ex...

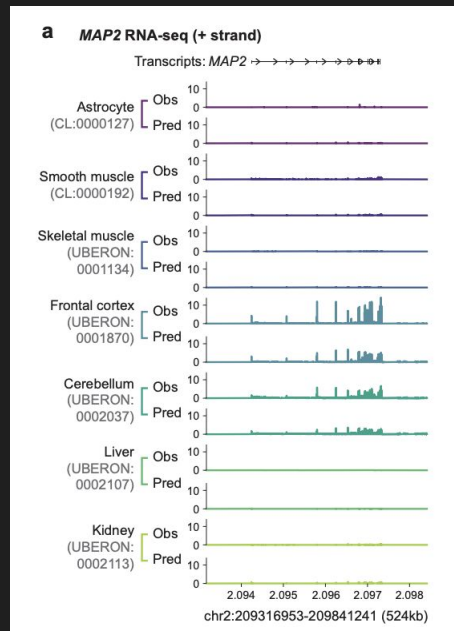
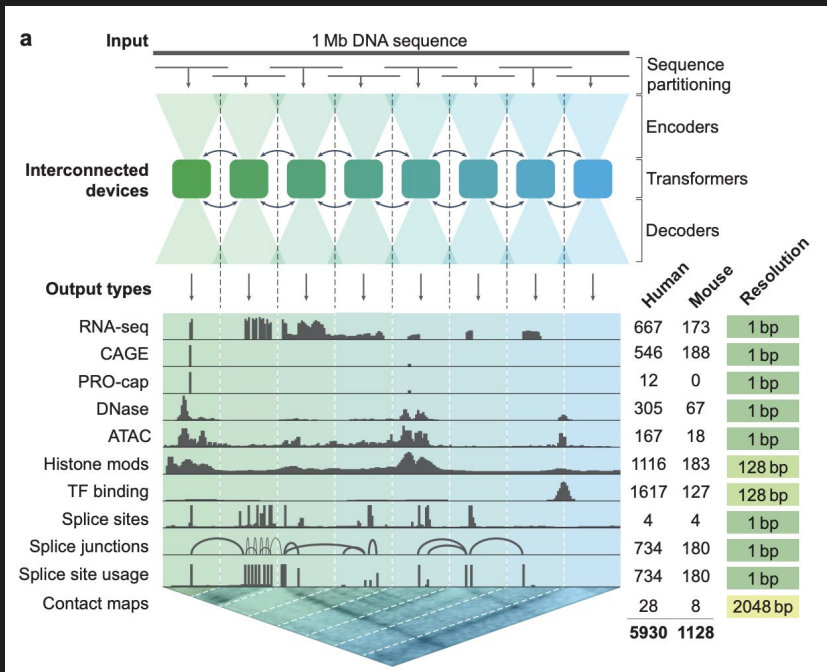
Discard

Save as draft

Submit

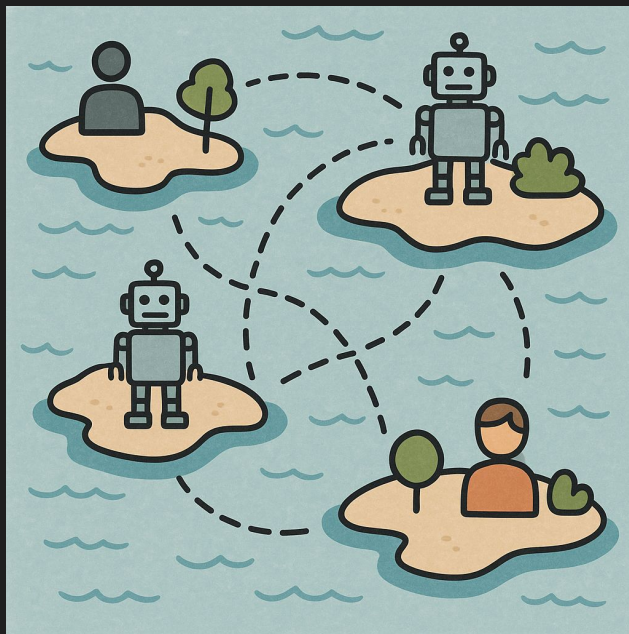
5

Ontologies powers AI



Avsec, Žiga, et al. "AlphaGenome: advancing regulatory variant effect prediction with a unified DNA sequence model." *bioRxiv* (2025): 2025-06.

Agentic AI can power ontology development



Semantic Social Coding Agents a framework for humans and AI collaborating together

Thank you

Monarch Initiative / OBO Academy

- **Sabrina Toro (UNC)**
- **Nico Matentzoglou (Semanticly)**
- Nicole Vasilevsky (Monarch, CPATH)
- Melissa Haendel (UNC)

Agent development and agent tools

- **Harshad Hegde (LBNL)**
- **Justin Reese (LBNL)**
- **David Osumi-Sutherland (Sanger)**
- **Marcin Joachimiak (LBNL)**
- **Harry Caufield (LBNL)**
- Jim Balhoff (RENCI)
- Carlo Kroll (UNC & QMUL)
- Seth Carbon (LBNL)
- Patrick Kalita (LBNL)
- Marcin Joachimiak (LBNL)
- Kevin Schaper (UNC)
- Andrew Schmeder (LBNL)
- Sierra Moxon (LBNL)
- James Overton (Knocean)
- **Damien Goutte-Gattat (FlyBase Cambridge)**

Uberon and CL testing

- David Osumi-Sutherland
- Chris Mungall

Mondo testing

- Sabrina Toro (UNC)
- Nicole Vasilevsky (Critical Path Institute)
- Yousif Schwetar (UNC)
- Katie Mullen (UNC)
- **Trish Whetzel (UNC)**

GO testing

- Pascale Gaudet (SIB)
- Stephen Marygold (FlyBase Cambridge)
- Val Wood (PomBase Cambridge)
- Raymond Lee (WormBase CalTech)
- Edith Wong (SGD Stanford)

METPO testing

- Mark Miller (LBNL)
- Marcin Joachimiak (LBNL)

MatKG testing

- David Abramov (LBNL)
- Chuck Parker (LBNL)
- Justin Reese (LBNL)
- Alex Hexemer (LBNL)

BioEpic ontology testing

- Harry Caufield (LBNL)
- Jinyun Tang (LBNL)
- Joan Damerow (LBNL)

Join us for a tutorial
at ICBO 2025
Nov. 9-11, 2025

(fully virtual, free
registration)

<https://ai4curation.github.io/aidocs/>

NHGRI Monarch Initiative
NHGRI Phenomics First
NHGRI Gene Ontology
LBNL LDRD (Contextualizer,
Fair2Wise)

Conclusions

- AI needs Ontologies needs AI
- Agentic AI is a major paradigm shift
- MCP all of the things!
- Ontology maintenance is hard... but agentic AI might help!
- OntoEval/SCRIBE is a step towards systematic evaluation...
 - ...but challenges remain

How do we evaluate?

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```

Edges unique to manually curated

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosolpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosolpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosolpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosolpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosolpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosolpinx
```


Metadata uniquely added by agent

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```


Differences in synonym handling

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```

Minor changes in wording

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberontology/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```

Changes in wording of definitions

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```

Metadiff: strict scoring based on diff of diffs

Curator

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "A serous layer that is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine (fallopian) tube to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [doi:10.1016/B978-0-12-801238-3.64401-X, PMID:29763118, Wikipedia:Mesosalpinx]
+synonym: "fallopian tube mesosalpinx" EXACT []
+synonym: "mesentery of fallopian tube" EXACT [PMID:29763118]
+synonym: "mesentery of uterine tube" EXACT [PMID:29763118]
+synonym: "serous layer of fallopian tube" EXACT [doi:10.1016/B978-0-12-801238-3.64401-X]
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
+is_a: UBERON:0000042 ! serous membrane
is_a: UBERON:0002095 ! mesentery
```

Agent

```
[Term]
id: UBERON:0012331
name: mesosalpinx
-def: "The mesosalpinx is part of the lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the uterine tube to the level of the ovary." [Wikipedia:Mesosalpinx]
+def: "The mesosalpinx is part of the peritoneal lining of the abdominal cavity in higher vertebrates, specifically the portion of the broad ligament that stretches from the fallopian tube (uterine tube) to the ovary. It is made up of epithelial cells of mesothelium and a mesh of serous membrane lining." [Wikipedia:Mesosalpinx, PMID:29763118]
+synonym: "serous layer of the fallopian tube" EXACT []
+term_tracker_item: "https://github.com/obophenotype/uberont/issues/3411" xsd:anyURI
+created_by: REDACTED
xref: FMA:19808
xref: SCTID:279900000
xref: Wikipedia:Mesosalpinx
```

F1 Score: 0.05

Agent made some different decisions that were arguably also valid

LLM as Judge

[Term]

id: UBERON:00

name: mesosal

-def: "The mes
specifically t
level of the o

+def: "A serou
er vertebrates
ine (fallopian
esh of serous

ipedia:Mesosal

+synonym: "fal

+synonym: "mes

+synonym: "mes

+synonym: "ser

xref: FMA:198

xref: SCTID:2

xref: Wikiped

+is_a: UBERON:

is_a: UBERON:

comments: *The primary difference between the two changes is the comprehensiveness in addressing the issue. The left change is more thorough in adding relationships and aligning with the discussed need for taxonomic generality.*

confidence: 0.8

difficulty: 0.6

issue_clarity: 0.8

left_evaluation:

evaluation: *The left change addresses the issue by updating the definition of 'mesosalpinx' to include more detailed anatomical information as requested. It adds synonyms that align with the suggestions and provides new relationships like 'is_a' serous membrane and 'surrounds' fallopian tube. The changes are consistent with the need for taxonomic generalization and acknowledge the complexity of the issue discussed in the comments.*

instruction_following_score: 0.9

overall_score: 0.9

logical_consistency: 0.7

right_evaluation:

evaluation: *The right change updates the definition of 'mesosalpinx' to include the requested additional information about its structure but makes fewer changes overall compared to the left. It adds a synonym but does not add new relationships such as 'is_a' or 'surrounds'. While it includes a term tracker item, it does not fully incorporate the structural complexity and relational aspects described in the issue and comments.*

instruction_following_score: 0.8

missing_changes:

- Missing 'is_a' link to serous membrane as seen in left change
- Missing 'surrounds' relationship to fallopian tube

overall_score: 0.7

score_diff: 0.2

similarity: 0.6

suggestions_for_users: *The issue could be worded more clearly by specifying the taxonomic context for the terms like 'uterine tube', 'fallopian tube', and 'oviduct'. Clarifying whether these terms are used interchangeably or have distinct uses in particular species could improve clarity.*

were arguably also valid

LLM as Judge: Score difference

[Term]

id: UBERON:00

name: mesosal

-def: "The mes
specifically t
level of the o

+def: "A serou
er vertebrates
ine (fallopian
esh of serous

ipedia:Mesosal

+synonym: "fal

+synonym: "mes

+synonym: "mes

+synonym: "ser

xref: FMA:198

xref: SCTID:2

xref: Wikiped

+is_a: UBERON:

is_a: UBERON:

comments: *The primary difference between the two changes is the comprehensiveness in addressing the issue. The left change is more thorough in adding relationships and aligning with the discussed need for taxonomic generality.*

confidence: 0.8

difficulty: 0.6

issue_clarity: 0.8

left_evaluation:

evaluation: *The left change addresses the issue by updating the definition of 'mesosalpinx' to include more detailed anatomical information as requested. It adds synonyms that align with the suggestions and provides new relationships like 'is_a' serous membrane and 'surrounds' fallopian tube. The changes are consistent with the need for taxonomic generalization and acknowledge the complexity of the issue discussed in the comments.*

instruction_following_score: 0.9

overall_score: 0.9

logical_consistency: 0.7

right_evaluation:

evaluation: *The right change updates the definition of 'mesosalpinx' to include the requested additional information about its structure but makes fewer changes overall compared to the left. It adds a synonym but does not add new relationships such as 'is_a' or 'surrounds'. While it includes a term tracker item, it does not fully incorporate the structural complexity and relational aspects described in the issue and comments.*

instruction_following_score: 0.8

missing_changes:

- Missing 'is_a' link to serous membrane as seen in left change

- Missing 'surrounds' relationship to fallopian tube

overall_score: 0.7

score_diff: 0.2

similarity: 0.6

suggestions_for_users: *The issue could be worded more clearly by specifying the taxonomic context for the terms like 'uterine tube', 'fallopian tube', and 'oviduct'. Clarifying whether these terms are used interchangeably or have distinct uses in particular species could improve clarity.*

er vertebrates,
ne tube to the

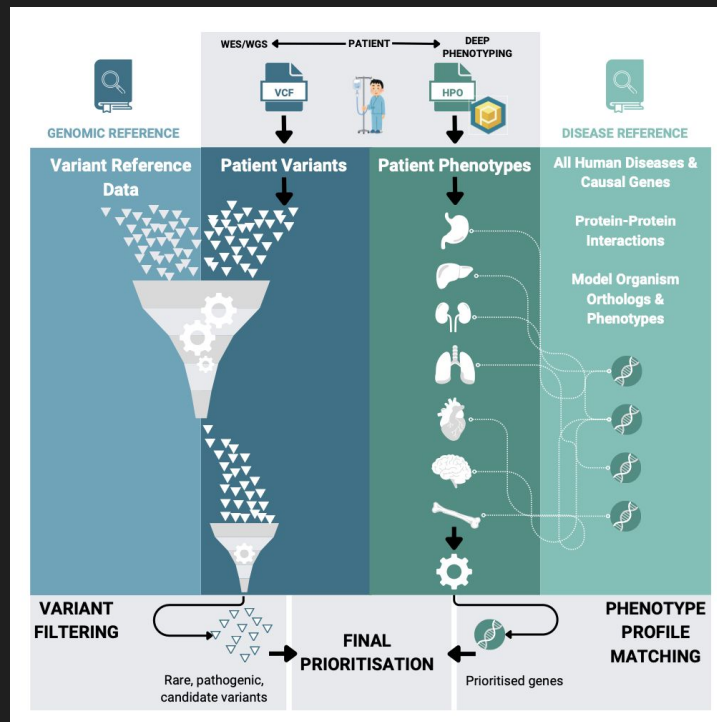
ity in higher v
om the fallopia
thelium and a

anyURI

at

were arguably also valid

Knowledge representation has explanatory power, not just predictive



Exomiser: leverages cross-species **phenomic knowledge** in addition to genomic knowledge for variant prioritization

Used for clinical explanatory power in diagnoses



Knowledge Base and Ontology Development: Team Science

Domain Expert
Glue
Product Owner
Consensus Engineer
Ontology Engineer
Data or Knowledge Curator
Community User

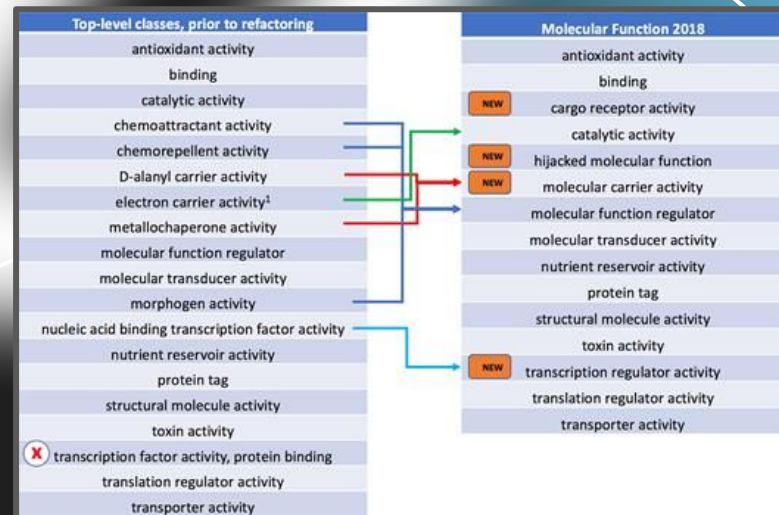
Changes

Most ontologies have a backlog of tickets and suffer from “technical debt”

Agentic AI can potentially help with various tasks

- Low level term requests
- Bulk changes
- Ontology refactoring
- Background research
- Issue management
- Checking for adherence to patterns and guidelines

But how do we put into practice?



refactoring

Open 653 Closed 20,097

Ticket triage and closing

The Ontology Development Kit: enabling social coding workflows for ontologies

Social Workflows: GitHub

Open discussion

...

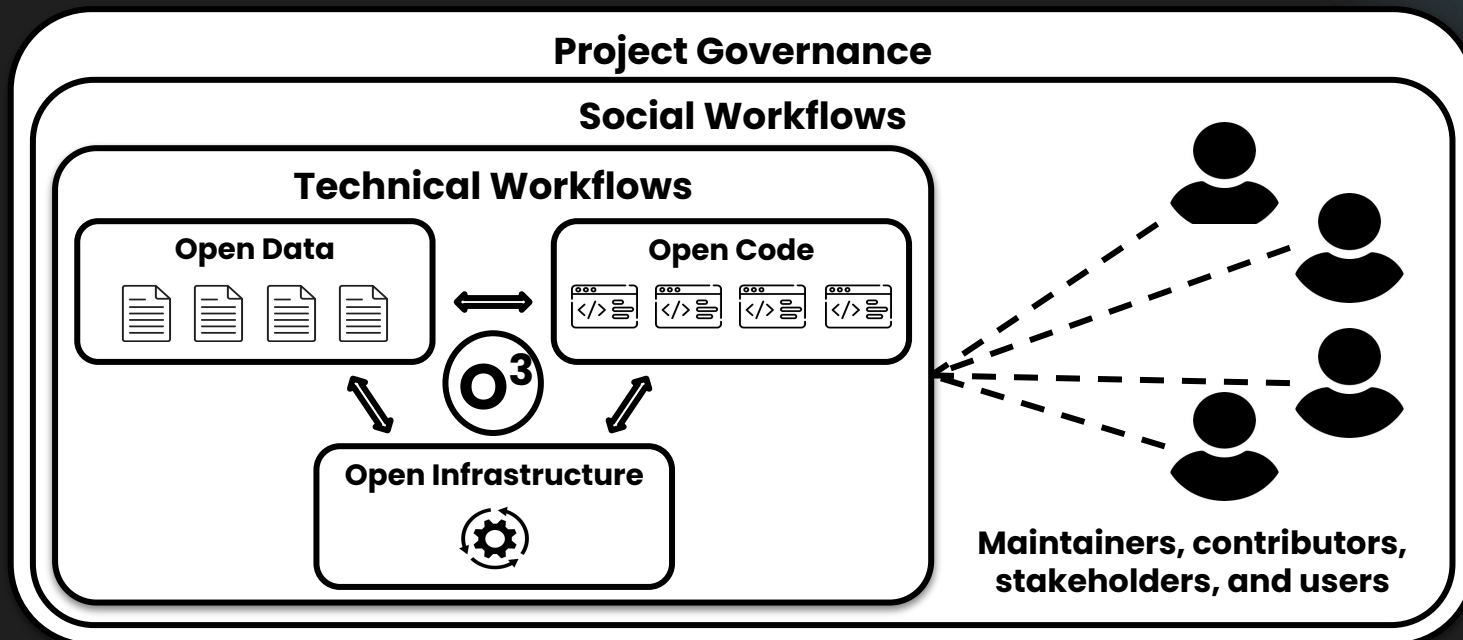
Open changes

...

Tool Integrations

...

The O3 guidelines: open collaborative development

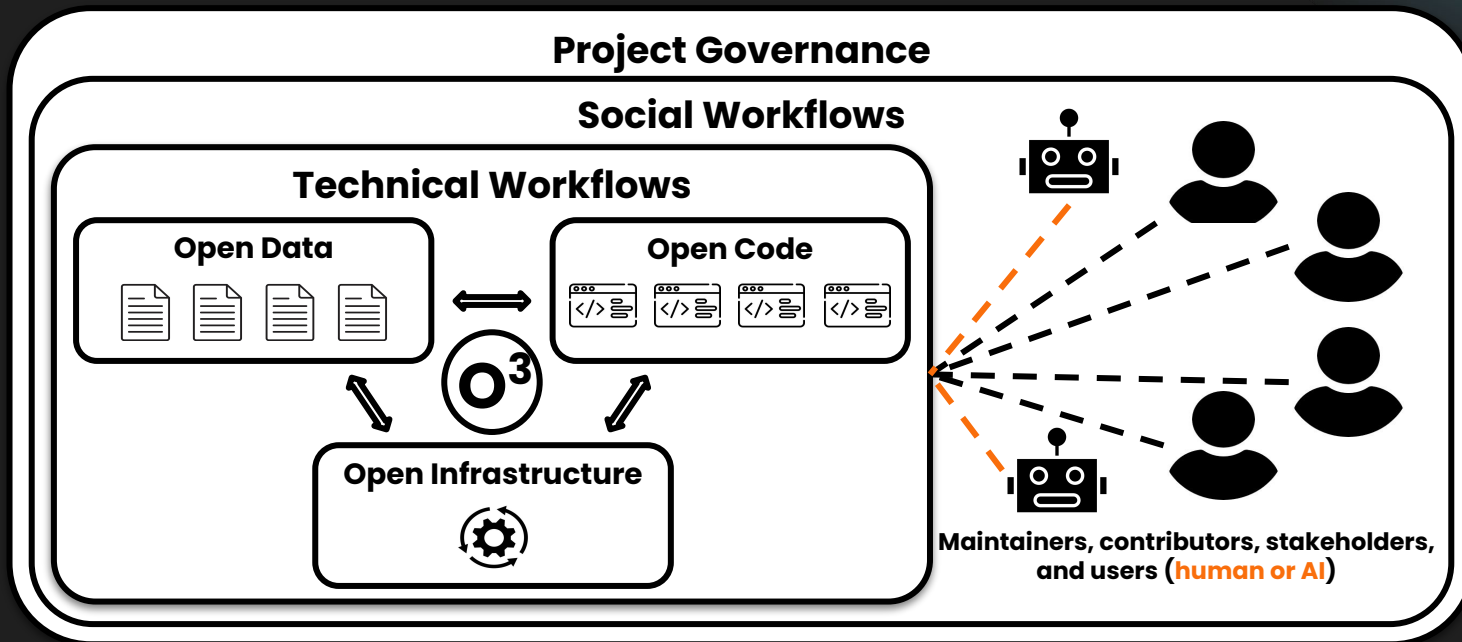


The O3 guidelines: open data, open code, and open infrastructure for sustainable curated scientific resources

Charles Tapley Hoyt & Benjamin M. Gyori

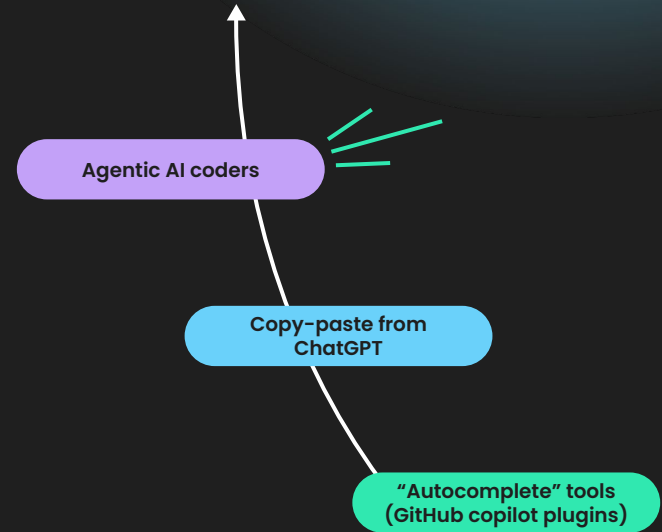
[Scientific Data](#) 11, Article number: 547 (2024) | [Cite this article](#)

The O3 guidelines: what happens when we add AI agents?

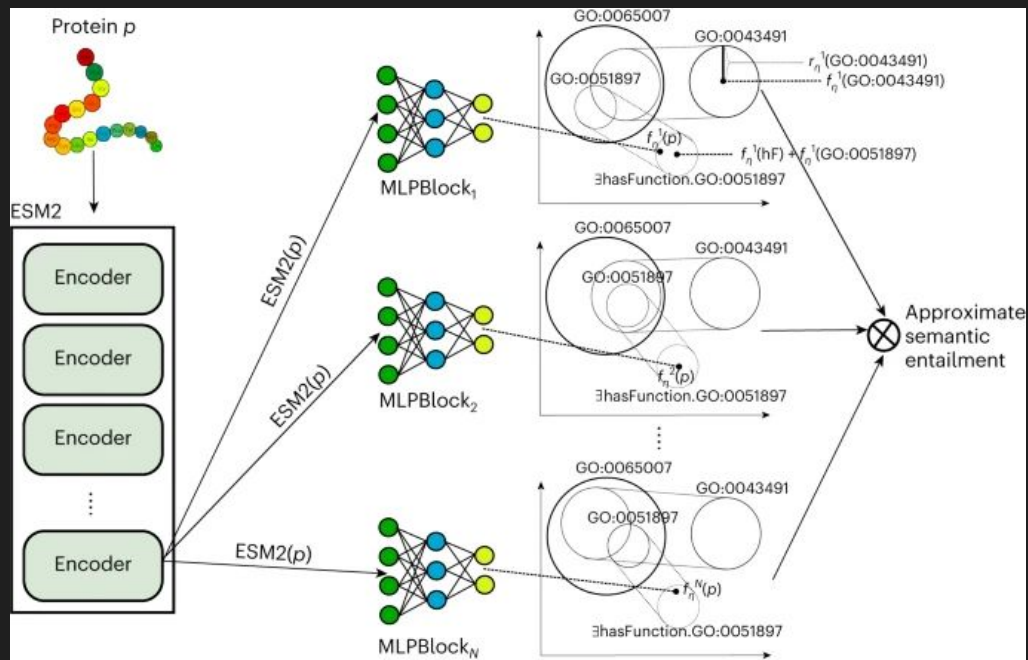


Note: the question is not **if** but **when**



GitHub as a collaborative environment for humans and agents



Knowledge representation powers AI

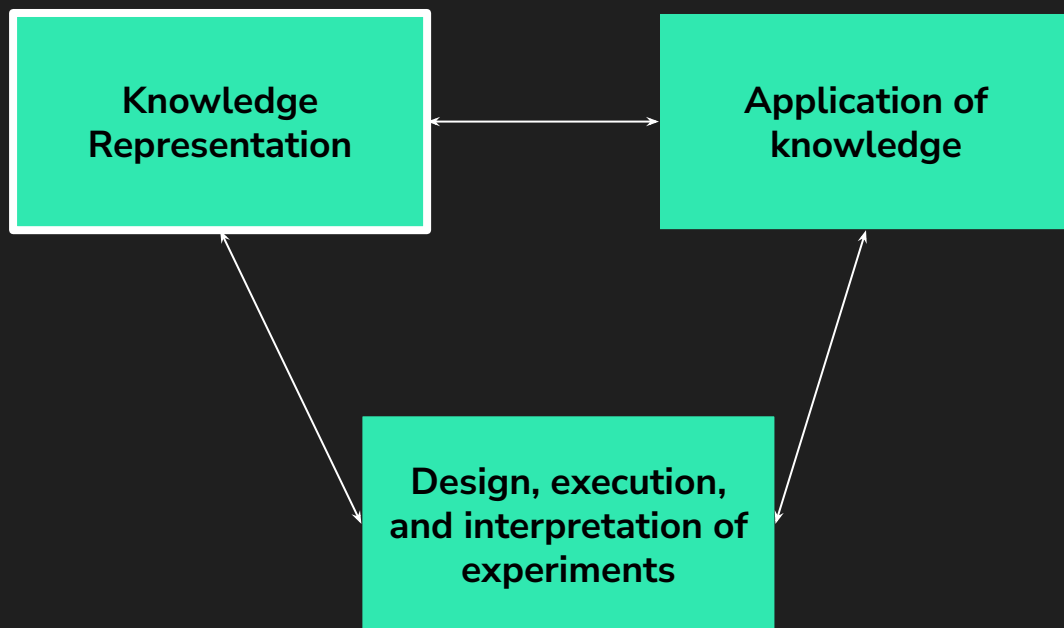


Protein function prediction as approximate semantic entailment

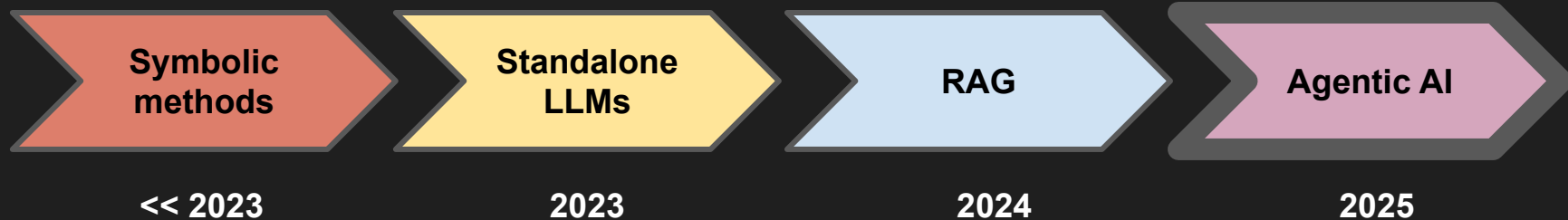
Maxat Kulmanov , Francisco J. Guzmán-Vega, Paula Duek Roggli, Lydie Lane, Stefan T. Arold & Robert Hoehndorf 

Nature Machine Intelligence **6**, 220–228 (2024) | [Cite this article](#)

How will AI impact this?



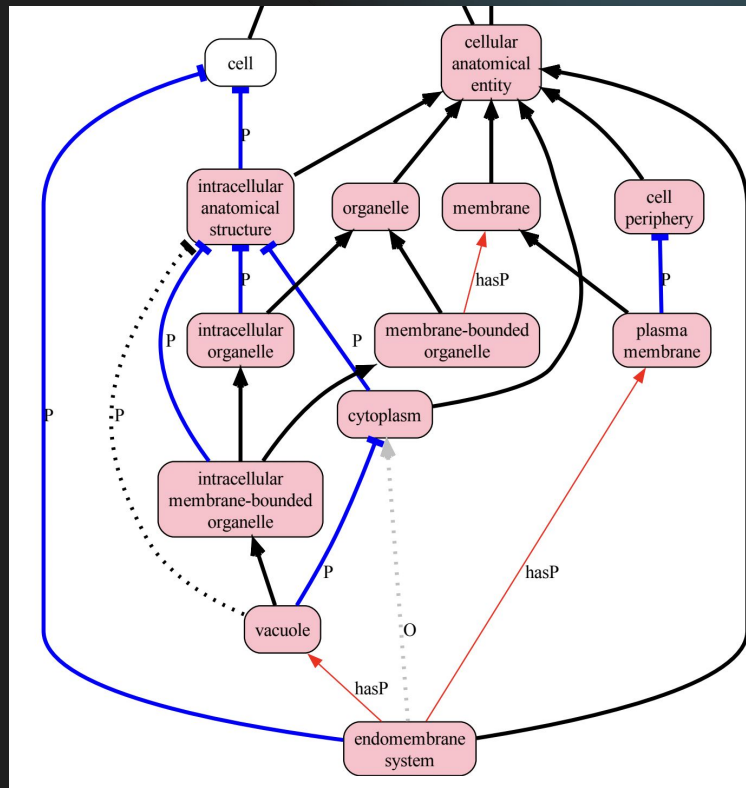
AI approaches to assist with knowledge representation: a short history



Symbolic approaches: description logics and reasoning

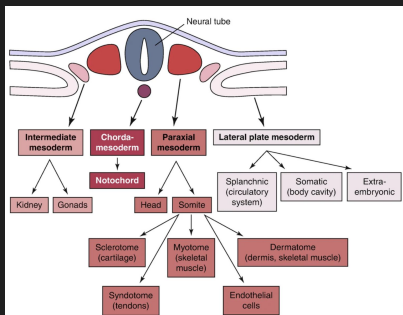
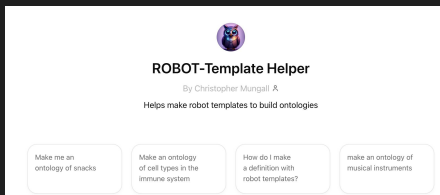
- Many ontologies rely logical axiomatization and **deductive reasoning** to automate many classification hierarchies
- ~50% of the *is-a* edges in GO are added by reasoners
- Critical for validation (e.g. taxon constraints)

**Symbolic
methods**



Balhoff, J.P. and Mungall, C.J., 2024. Whelk: An OWL EL+ RL Reasoner Enabling New Use Cases. Transactions on Graph Data and Knowledge, 2(2), pp.7-1.

Standalone LLMs



Standalone LLMs

Here's the CSV format robot template:

CSV

```

identifier,name,parent,develops_into
ID,LABEL,SC %,A 'develops into'
EMBRYO:001,Neural tube,,Chorda-mesoderm|Paraxial mesoderm
EMBRYO:002,Intermediate mesoderm,EMBRYO:001,Kidney|Gonads
EMBRYO:003,Chorda-mesoderm,EMBRYO:001,Notocord
EMBRYO:004,Paraxial mesoderm,EMBRYO:001,Head|Somite
EMBRYO:005,Lateral plate mesoderm,EMBRYO:001,Splanchnic
EMBRYO:006,Notocord,EMBRYO:003,
EMBRYO:007,Head,EMBRYO:004,
EMBRYO:008,Somite,EMBRYO:004,Sclerotome (cartilage)|Myotome
EMBRYO:009,Splanchnic (circulatory system),EMBRYO:005,
EMBRYO:010,Somatic (body cavity),EMBRYO:005,
EMBRYO:011,Extra-embryonic,EMBRYO:005,
EMBRYO:012,Kidney,EMBRYO:002,
EMBRYO:013,Gonads,EMBRYO:002,
EMBRYO:014,Sclerotome (cartilage),EMBRYO:008,

```



Make an ontology from this diagram

ChatGPT can make

<https://bit.ly/robot-gpt>

Standalone LLMs: ChatGPT-as-Oracle is an anti-pattern

- Works for toy examples or small ontologies
- Unable to hold mid-size to large ontologies in memory
- Does not fit with existing workflows
- Prone to hallucination

**Standalone
LLMs**

Retrieval Augmented Generation (RAG): Augmenting LLMs with knowledge

Create a new Class

Name spikelet floral color

IRI http://purl.obolibrary.org/obo/TO_00002200

New entity options...

Cancel OK

Question

"Add term 'spikelet floral color'"

Annotations +

rdfs:label
spikelet floret color

definition
A spikelet floret morphology trait (TO:0000274) which is the color of a spikelet floret (PO:0009082).

database_cross_reference
Gramene:pankaj_jaiswal

Description: spikelet floret color

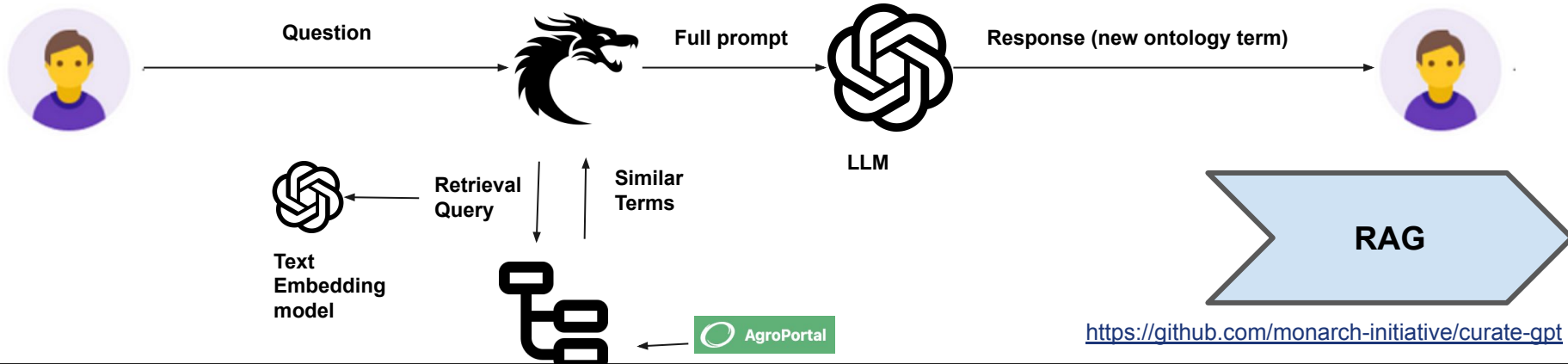
Equivalent To +
color
and ('characteristic of' some 'spikelet floret')

SubClass Of +
'flower color'
'spikelet floret morphology trait'

Definition
(text)

Logical
(OWL)
definition

Relationships



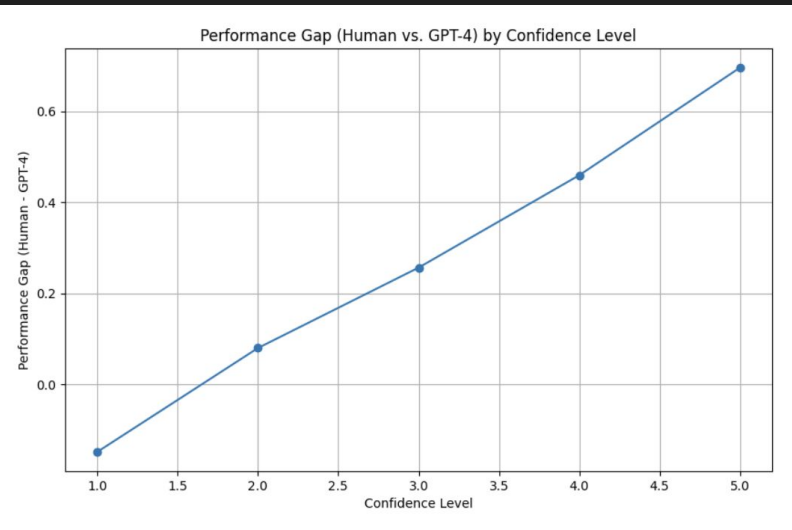
Evaluation of RAG on ontologies

Is-a link prediction task: High precision, poor recall

method	model	precision	recall	f1 score
DRAGON	gpt-4	0.889	0.44	0.588

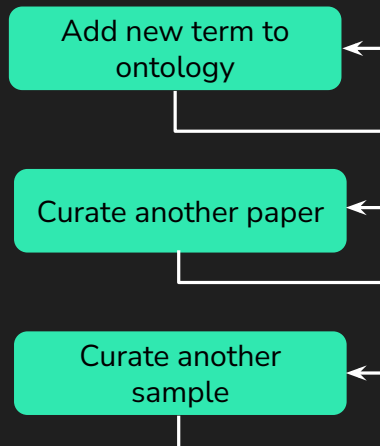
Definition task: AI attempts to “hoodwink” less experienced evaluators

Toro, Sabrina, et al. "Dynamic retrieval augmented generation of ontologies using artificial intelligence (DRAGON-AI)." *Journal of Biomedical Semantics* 15.1 (2024): 19.



Problem with first generation genAI approaches: Unsuitability for real-world knowledge representation tasks

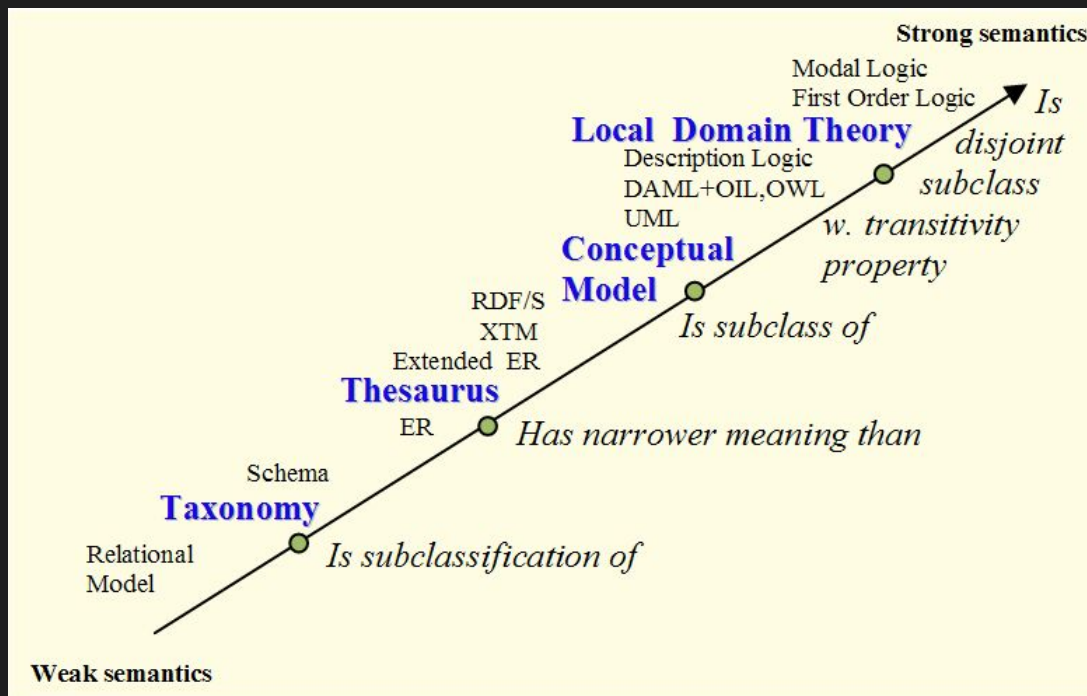
What AI can (sometimes) help with



What we **need** it to help with

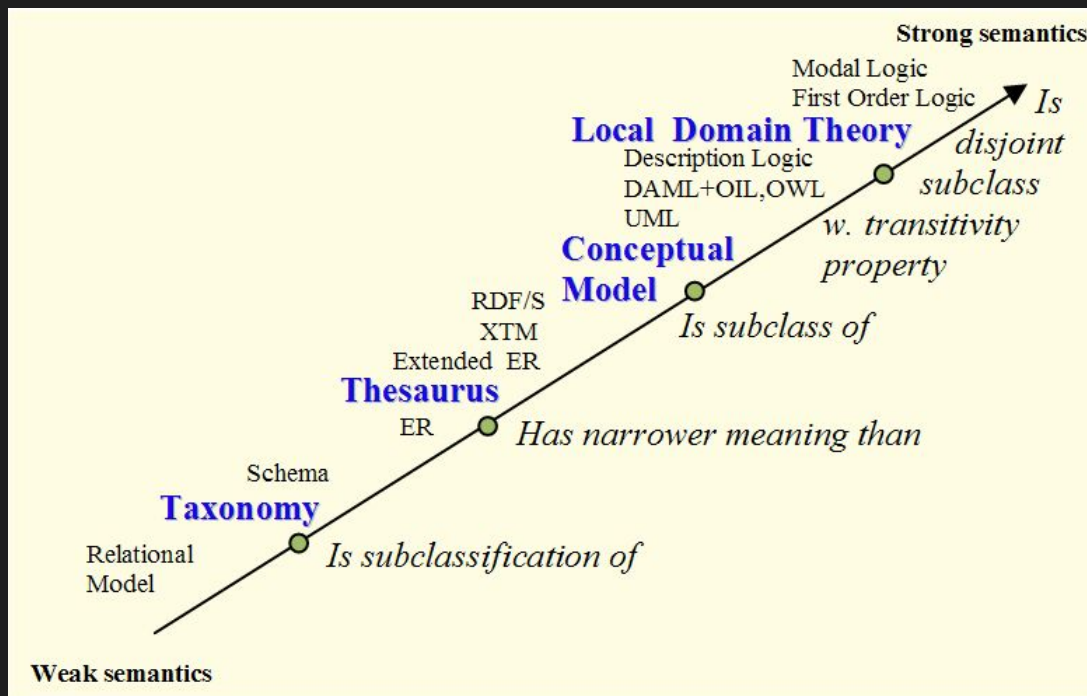


Semantic Ladder



Adapted from Daconta, Orbst, Smith 2003

Semantic Ladder: the next rung



Adapted from Daconta, Orbst, Smith 2003