Ontology Deconstruction and Recycling

Jose Antonio Vera Ramos, María del Mar Roldán, Sandro Hurtado,
Jesualdo Tomás Fernández-Breis, Stefan Schulz

GREEKC WP1 Workshop, Brussels, January 27, 2020

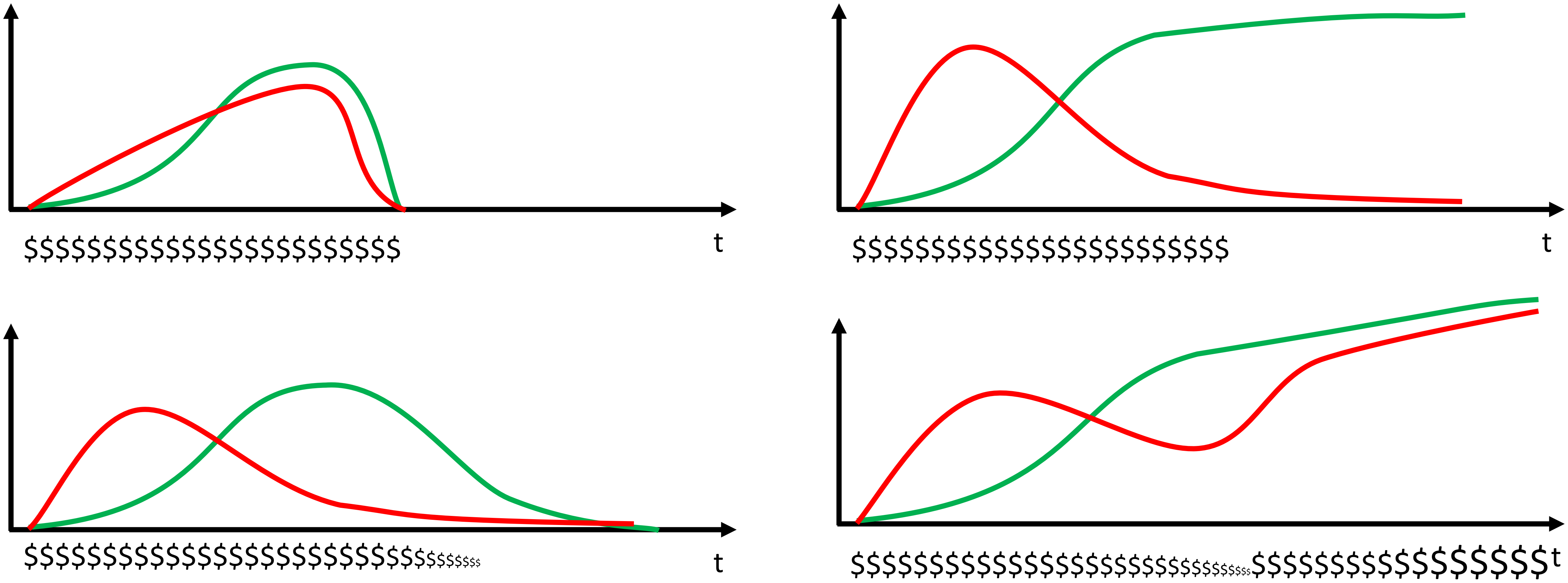
Development of generic method

- Rationale
  - "Grooming" the ecosystem of knowledge commons
  - Addressing obsolescence
  - Re-using content
  - Concentrating content in popular ontologies that are well-maintained
  - Preventing dead links
- Relevance for GREEKC
  - Addresses increasingly important problem with knowledge commons, particularly ontologies

Relevance for GRAO

- To be explored:
  - Ontology Deconstruction and Recycling methodology could be a starting point for an application ontology
  - Deconstructed ontology could serve as view on set of relevant ontologies
- Not in the case of GRAO
  - Different reasons: focus of original GRO different from what is needed now
  - GRAO therefore driven by GREEKC use cases, regardless of scope and granularity of original GRO
- Two separate tasks to be treated separately
  - 1. Deconstruction / recycling of existing ontologies
  - 2. Creation of a new application ontology
  - Overlaps / synergies between both tasks?

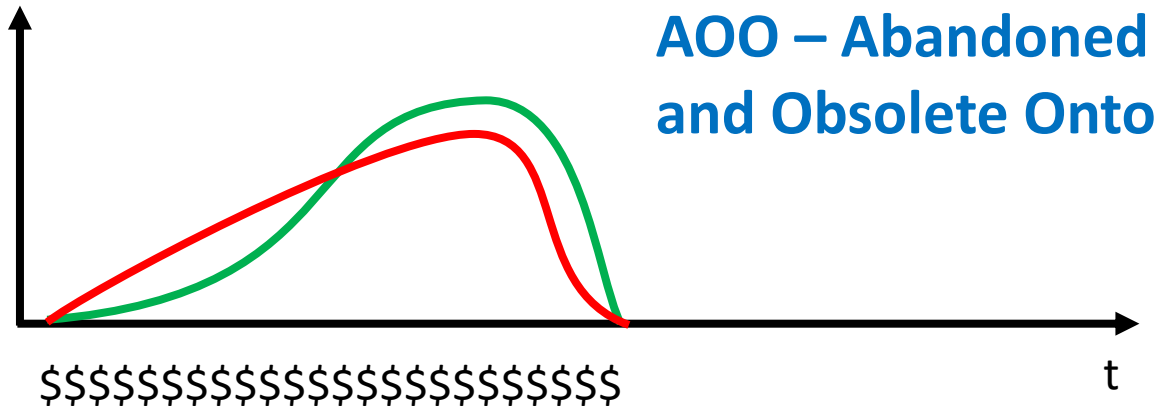
Knowledge commons life cycle variants

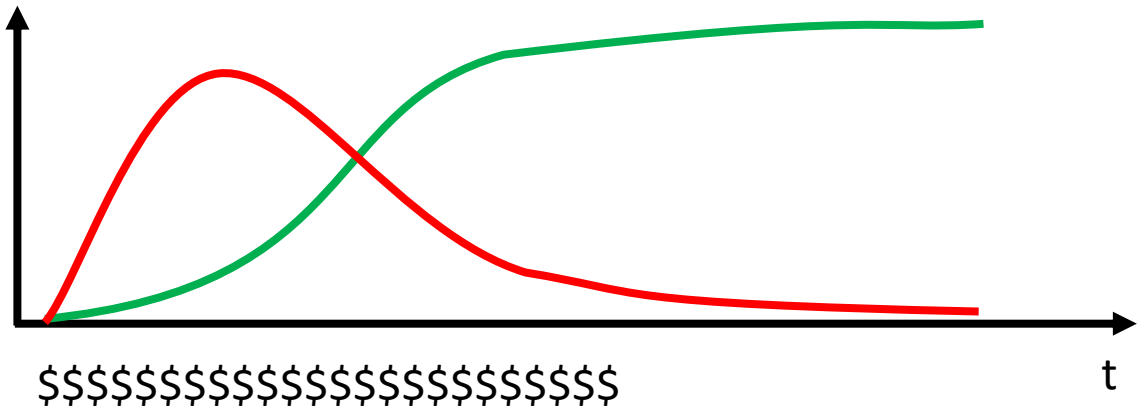


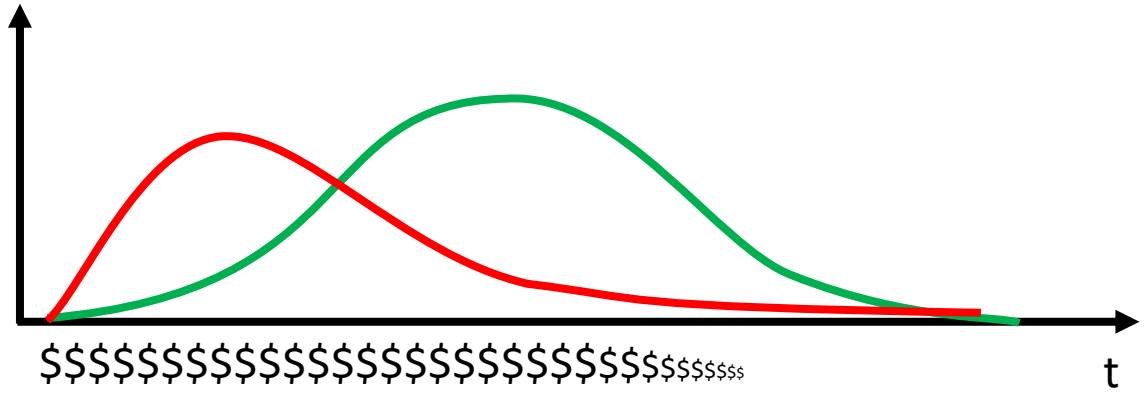
Maintenance

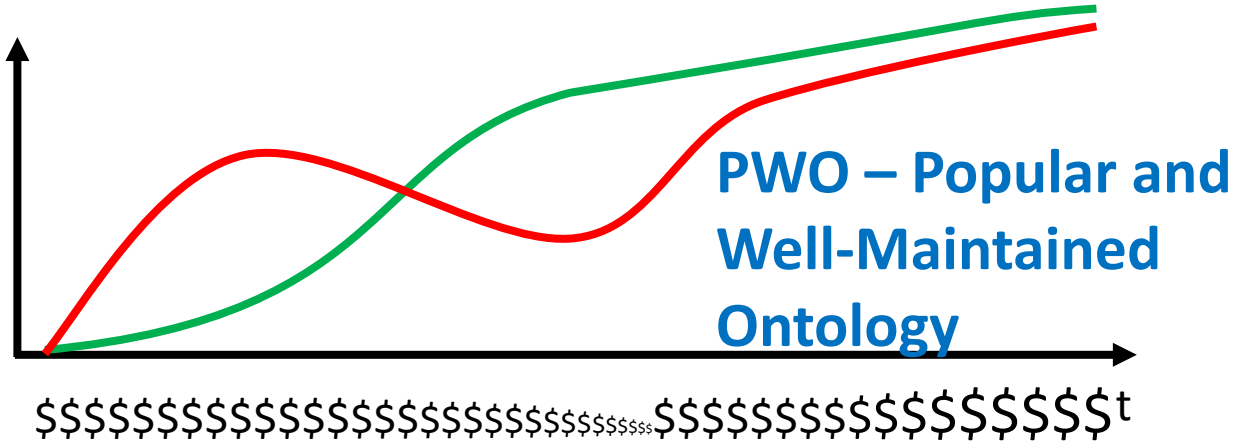
Use

Characterising publically available ontologies





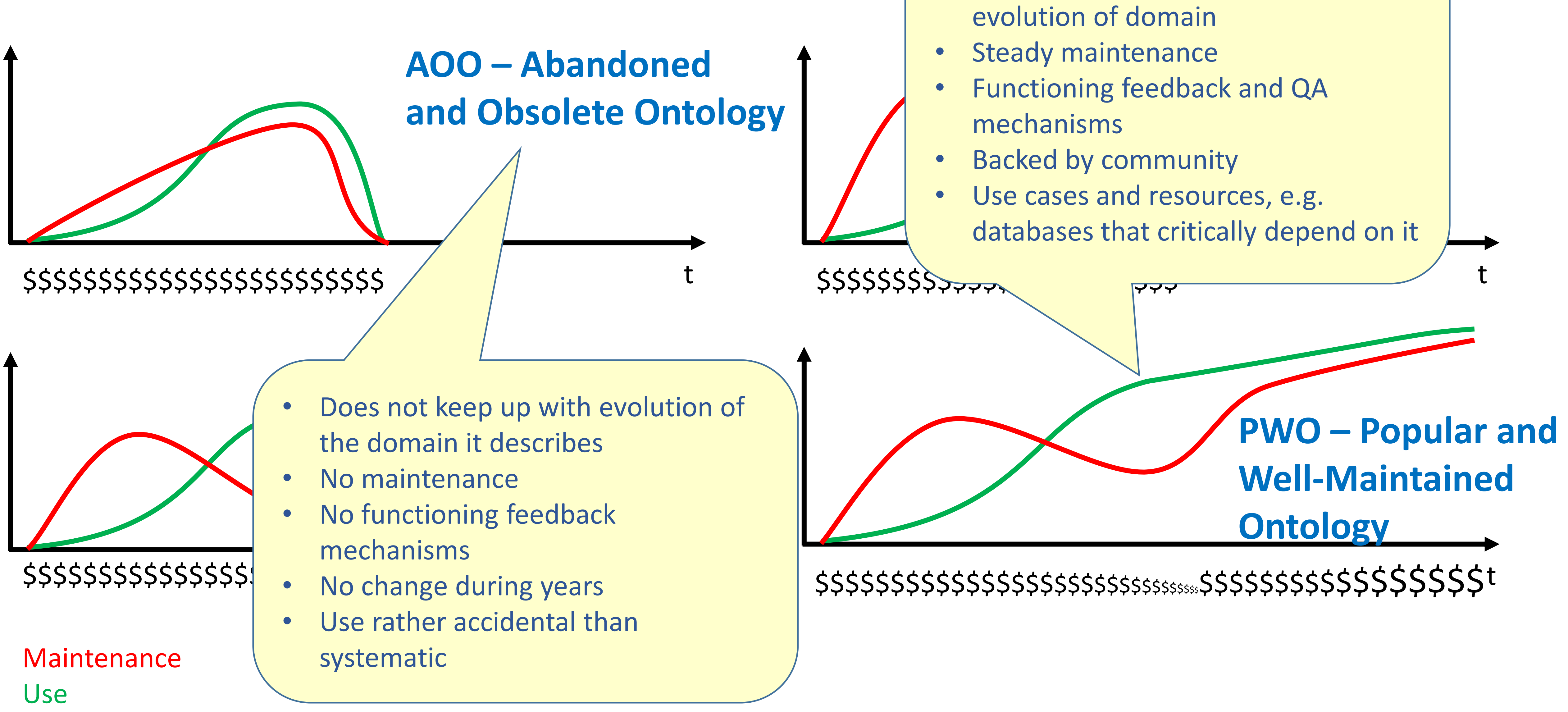




Maintenance

Use

Characterising publically avai



Why preserving (parts of) AOOs ? (abandoned & obsolete ontologies)

- Growing proportion due to permanence in public repositories
  - > 70% of Bioportal or even more?
- Still available as source for URIs
  - Referenced by other ontologies
  - Used in applications
  - Used in annotated resources
- Some content unique – not available in any other popular & well-maintained ontology (PWO)
- Some content redundant – represented in some other PWO
- Watch out – rarely complete semantic equivalence (re implicitly assumed meaning, textual and/or formal definitions)

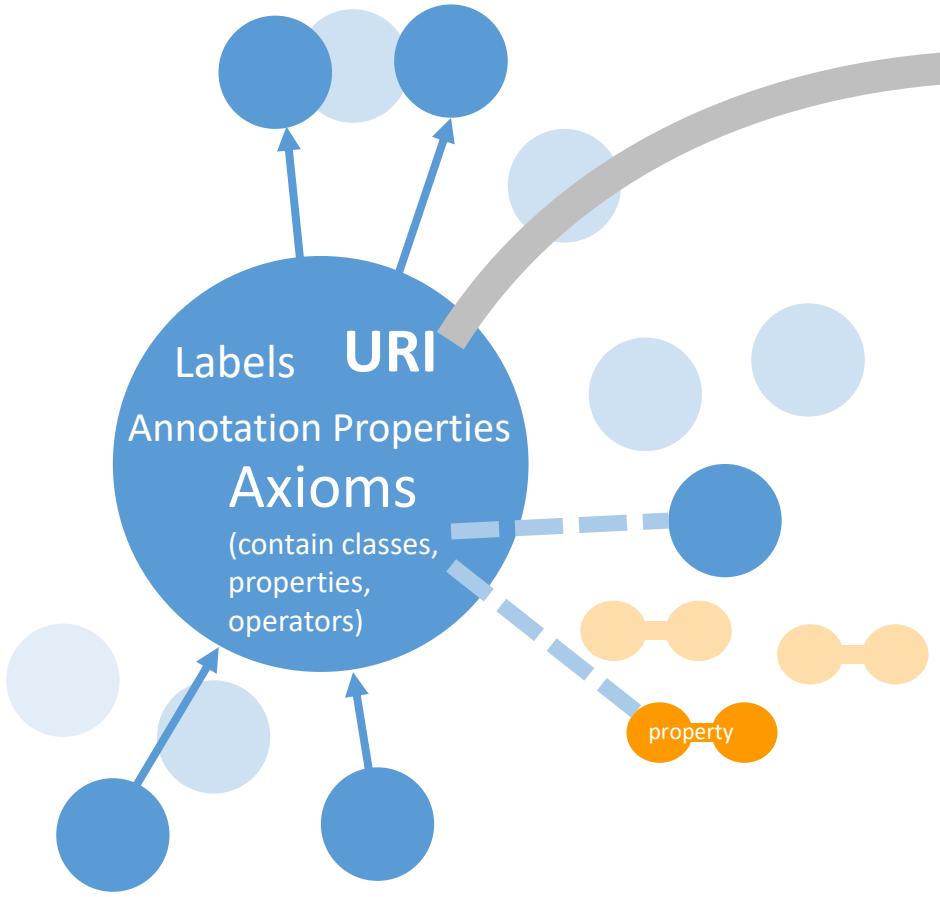
Proposal: grooming ontology ecosystem by controlled content deconstruction and recycling

- For a given AOO in a version y: $AOO_{k.y}$
  - Transformation:
    $AOO_{k.y}$ → $AOO_{k.z}$ for z = final and stable, linked to PWO content
- Goals:
  - Partly automatized process, with limited investment of manual work
  - Preservation of all identifiers (classes, properties) of $AOO_{k.y}$
  - Selection of related PWOs
  - URI Redirection to related PWOs
  - Content submission requests to related PWOs
  - Ideal: $AOO_{k.z}$ not more than a collection of external ontology URIs, with a minimal of meta-information

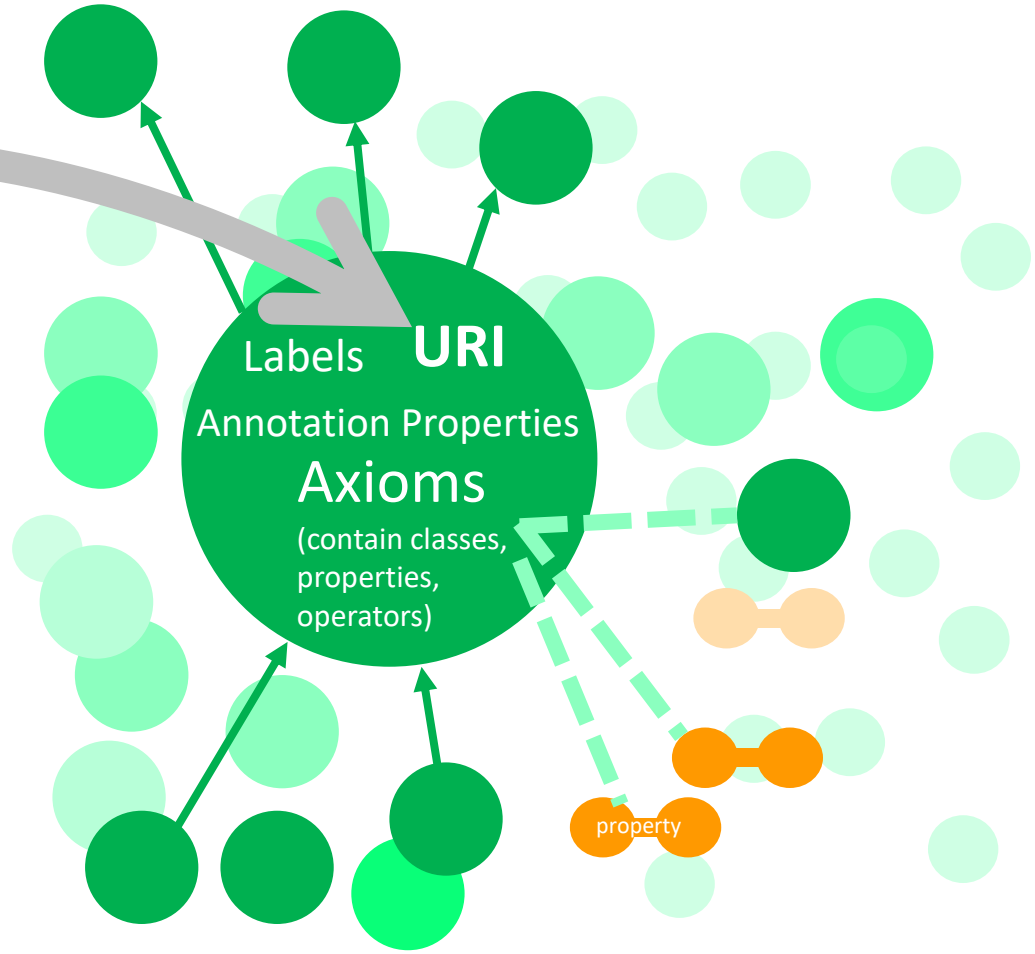
Selecting the ideal target ontologies

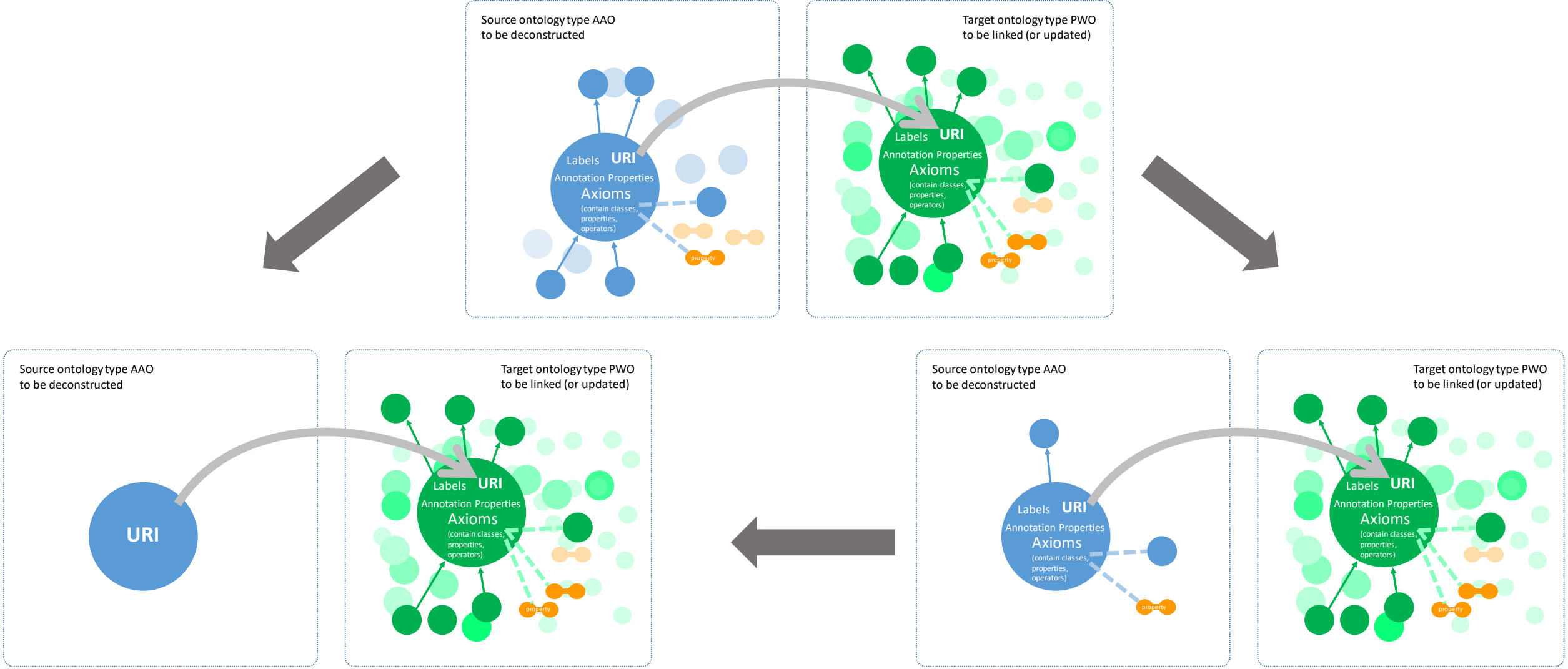
- Popular (= frequently used in other ontologies, applications and annotations)
- Well-maintained (community support)
- Created under a popular upper-level ontology
- Exclusion criteria
  - Artefacts that are not ontologies in a strict sense (not using OWL syntax and semantics)
  - Ontologies to which access and use restrictions apply (e.g. SNOMED CT)
  - Ontologies that do not considered that do not provide stable URIs
  - Application ontologies

Source ontology type AAO
to be deconstructed



Target ontology type PWO
to be linked (or updated)





Preservation of source URI only

1. Target class already exists that fully represents the meaning of the source URI
2. Equivalent target class has been created after content submission meaning of the source URI

Preservation of source URI, together with some information not present in the target ontology, e.g.

Axioms, metadata

Further details and GRO use case

RESEARCH

Deconstructing and recycling ontologies: A Gene Regulation ontology use case

José Antonio Vera-Ramos[1*†], María del Mar Roldán-García[2,3,4†], Sandro Hurtado[2,3], Jesualdo Tomás Fernández-Breis[5] and Stefan Schulz[1]

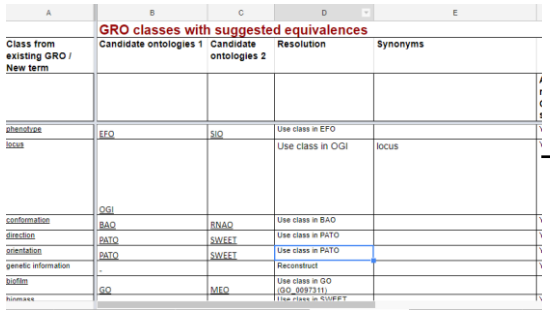
*Correspondence:
jose.vera-ramos@medunigraz.at
[1]Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerpl. 2, 8036 Graz, Austria
Full list of author information is available at the end of the article
†Equal contributor

Abstract

Background: Biomedical ontologies play an important role for biomedical data interoperability. To this end, hundreds of ontologies have been made available in ontology repositories. Some of these ontologies may have been useful in the past but are no longer maintained. Their content may be of interest for the community but it is potentially outdated and incomplete. Besides, content may overlap with other well-maintained ontologies. Updating such obsolete ontologies to be aligned with well-maintained ones would make them useful again. We address this problem by discussing the concept of ontology life cycles, by

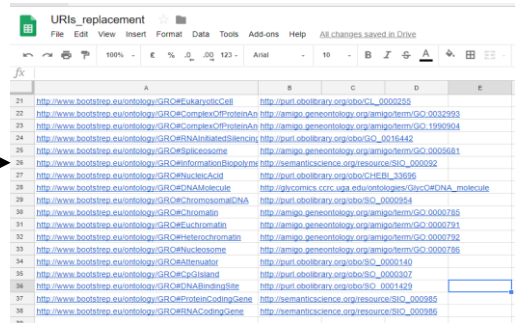
https://de.overleaf.com/9598645218ctfmpjgqtgzs

GRO recycling



GRO classes with suggested equivalences

1

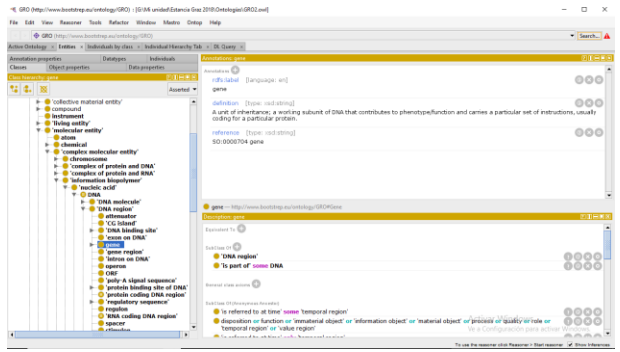


OLD_URI, NEW_URI

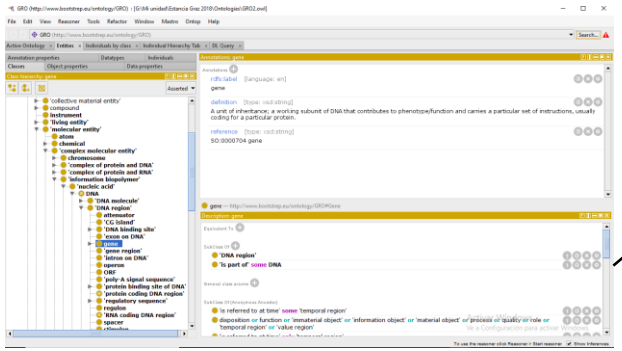
2

python

3



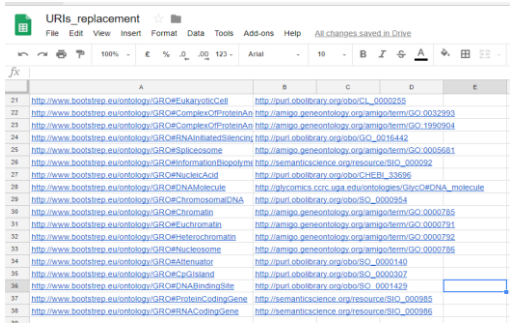
Recycled GRO



2

GRO

GRO recycling



OLD_URI, NEW_URI

New_URI

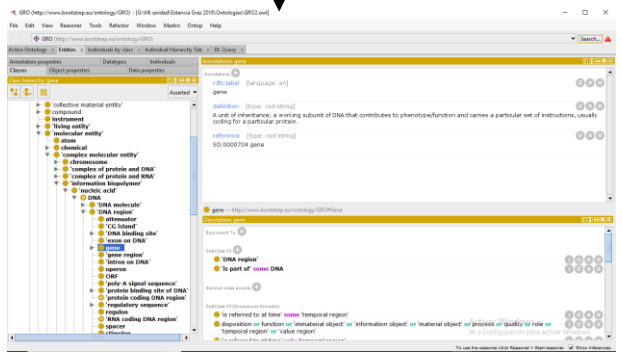
python

Bioportal API
SparqlWrapper[1]

label, definition, axioms

python

OWL API
owlready2[2]



Recycled GRO

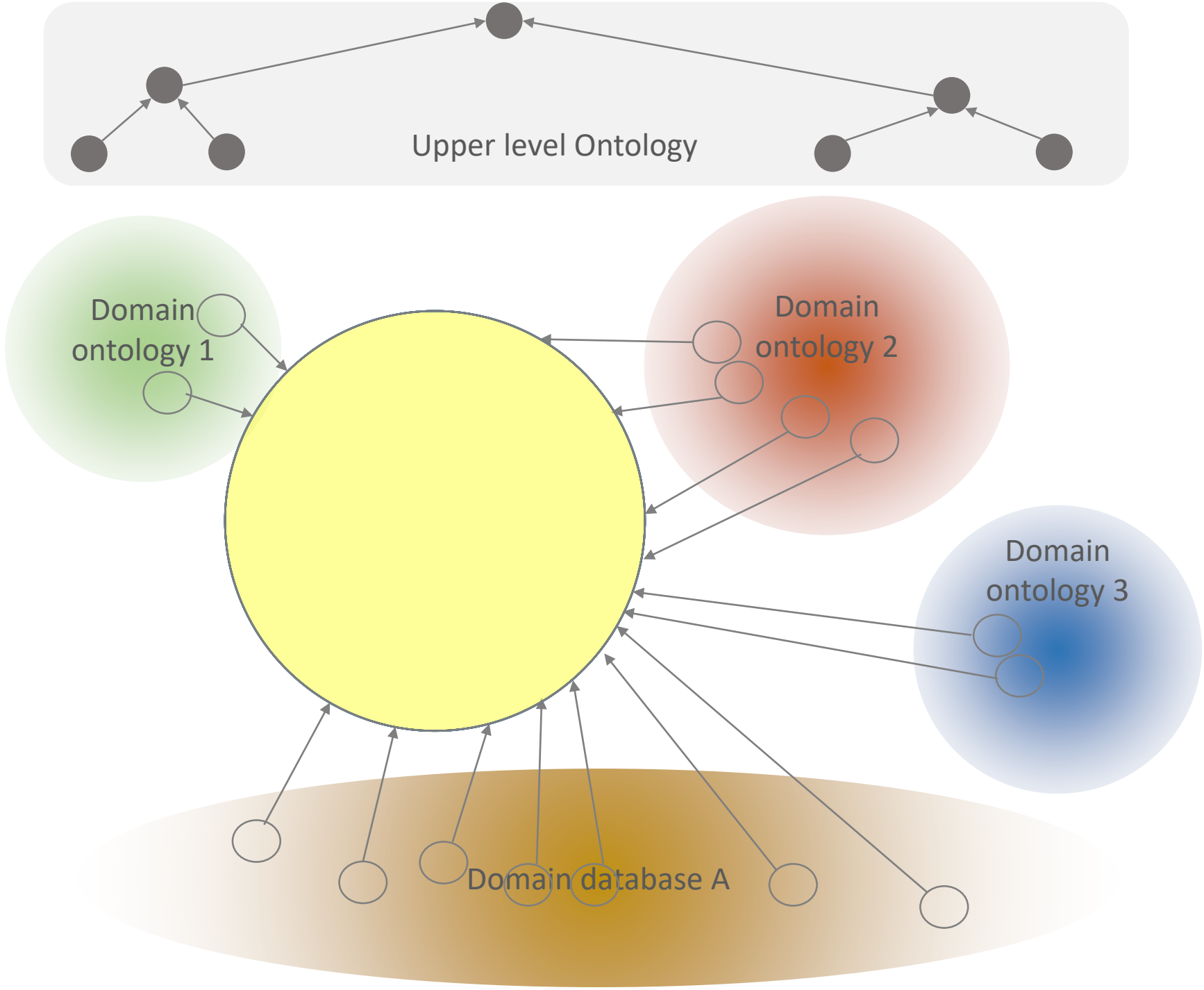
BioPortal

[1]https://rdflib.github.io/sparqlwrapper/

[2]Lamy JB. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artificial Intelligence In Medicine 2017;80:11-28

Open issues

- Extend workflow described for classes, for other OWL elements, such as object properties, datatype properties, annotation properties, axioms
- What is realistic? How do ontology creators react to content inclusion requests
- Which compromises are acceptable (between representation in source and target ontology)
- How can a "de novo" application ontology construction process (like GRAO) re-use (parts of) this methodology?



Labels

Annotation Properties
Axioms

(contain classes, properties, operators)