



Gene Regulation Ensemble
Effort for the Knowledge
Commons

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

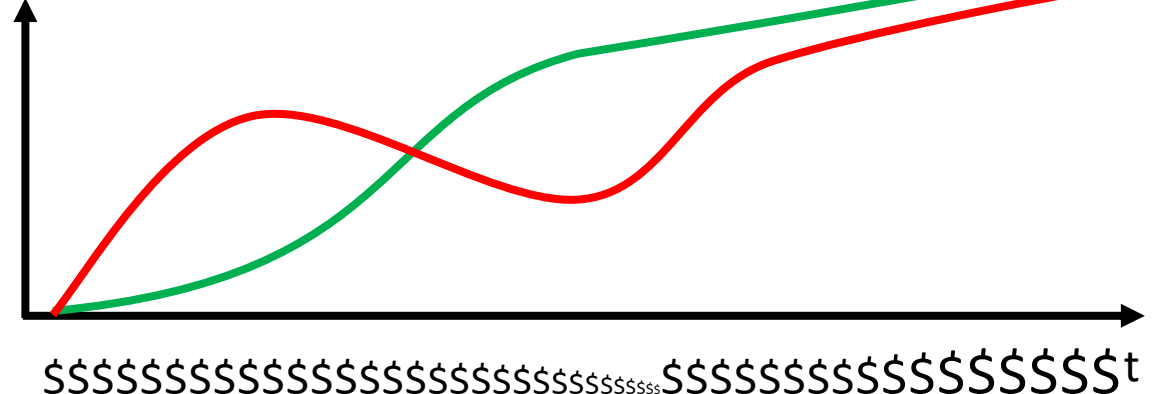
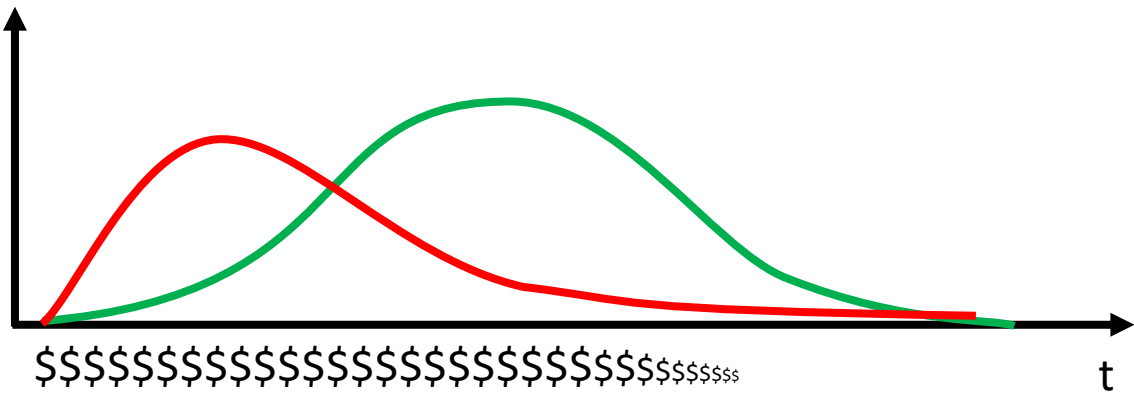
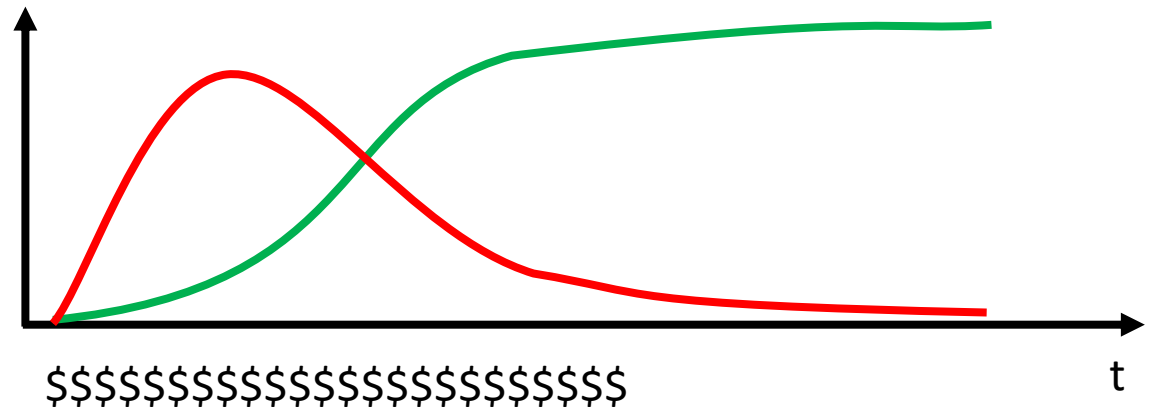
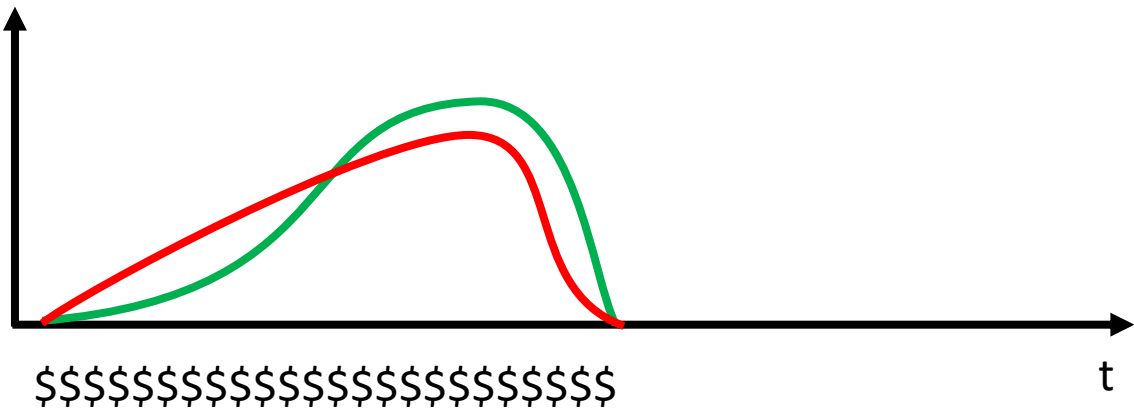
Development of generic method

- 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

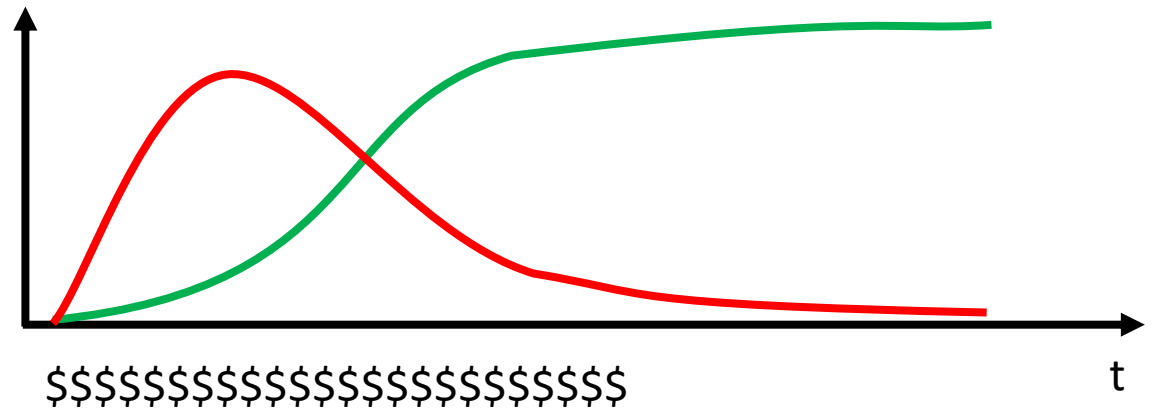
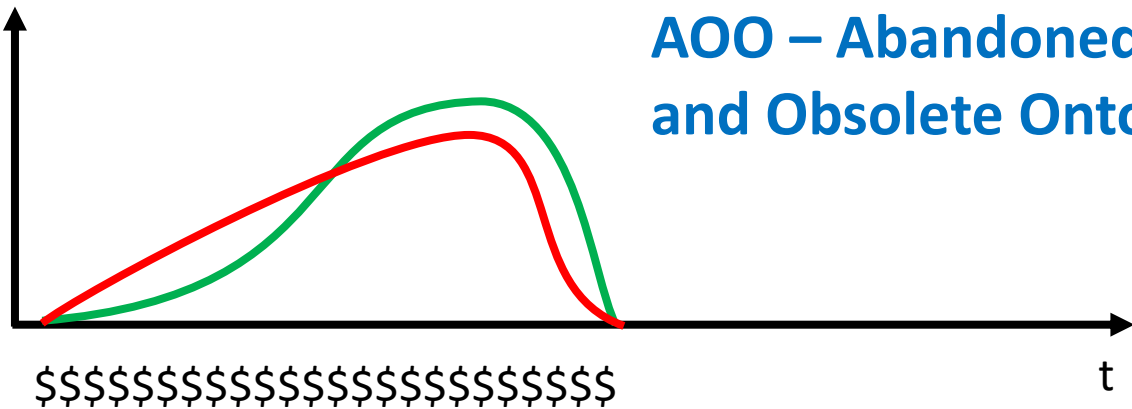


Maintenance

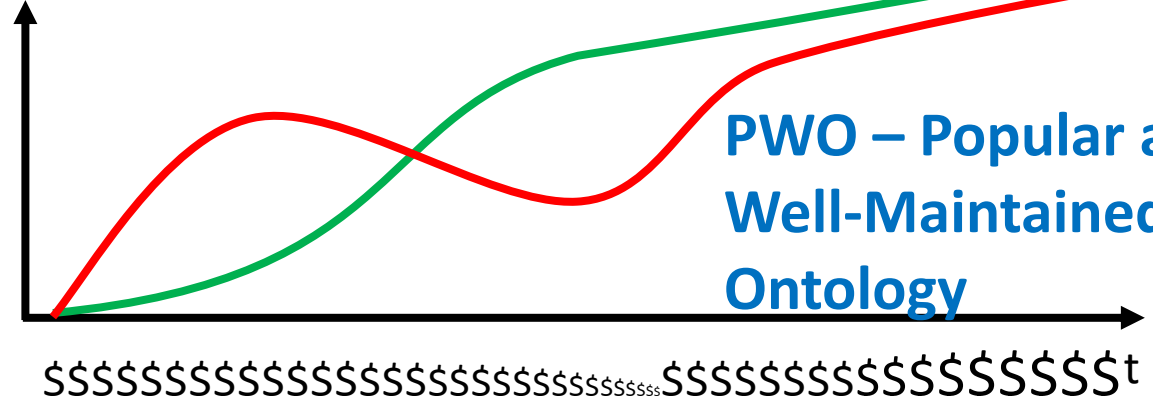
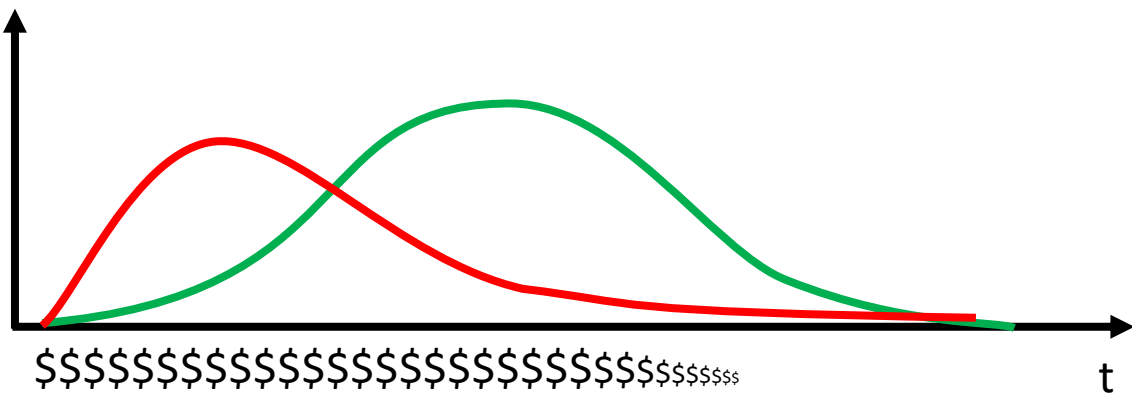
Use

Characterising publically available ontologies

AOO – Abandoned and Obsolete Ontology



PWO – Popular and Well-Maintained Ontology

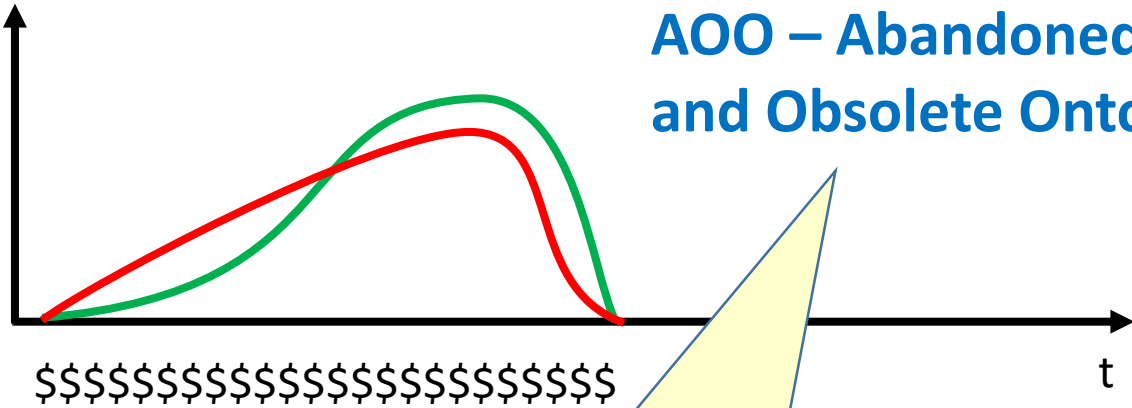


Maintenance

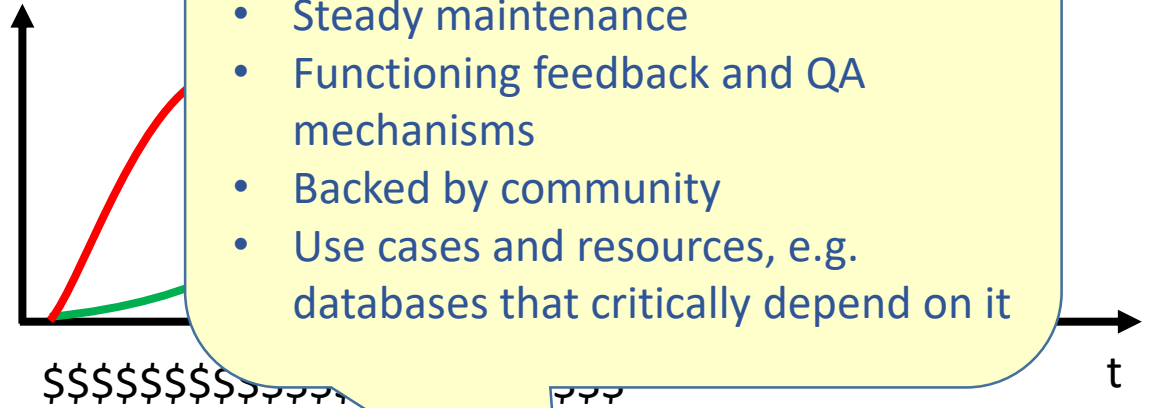
Use

Characterising publically available

AOO – Abandoned and Obsolete Ontology

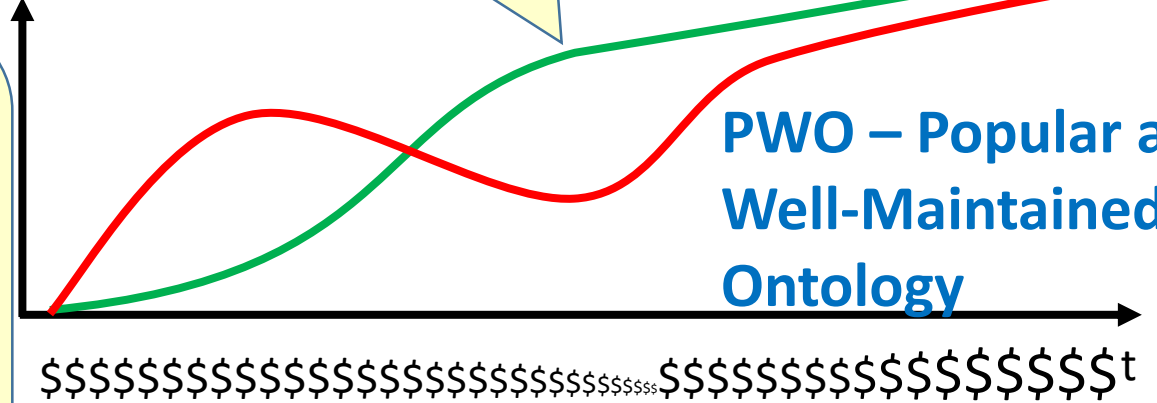
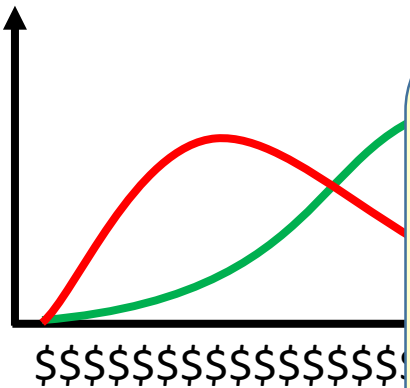


- Frequently updated according to evolution of domain
- Steady maintenance
- Functioning feedback and QA mechanisms
- Backed by community
- Use cases and resources, e.g. databases that critically depend on it



- Does not keep up with evolution of the domain it describes
- No maintenance
- No functioning feedback mechanisms
- No change during years
- Use rather accidental than systematic

PWO – Popular and Well-Maintained Ontology



Maintenance
Use

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)

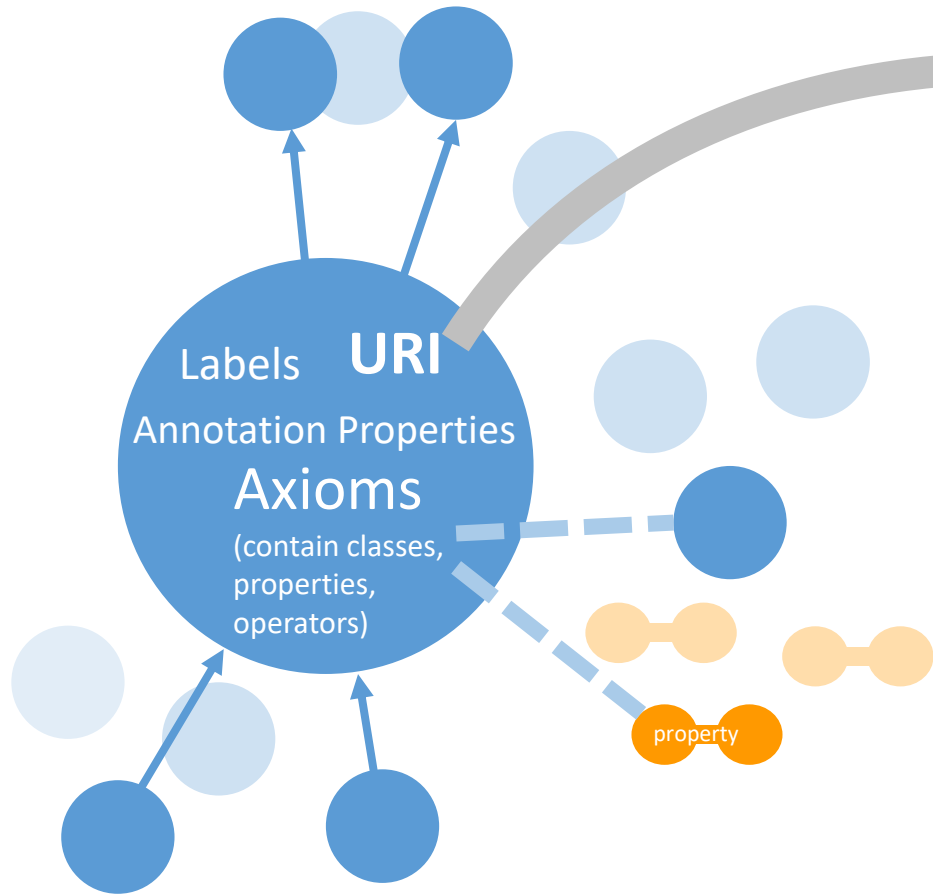
Proposal: grooming ontology ecosystem by controlled content deconstruction and recycling

- For a given AOO in a version y : $AOO_{k.y}$
 - Transformation:
 $AOO_{k.y} \rightarrow AOO_{k.z}$ for $z = \text{final and stable}$, 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

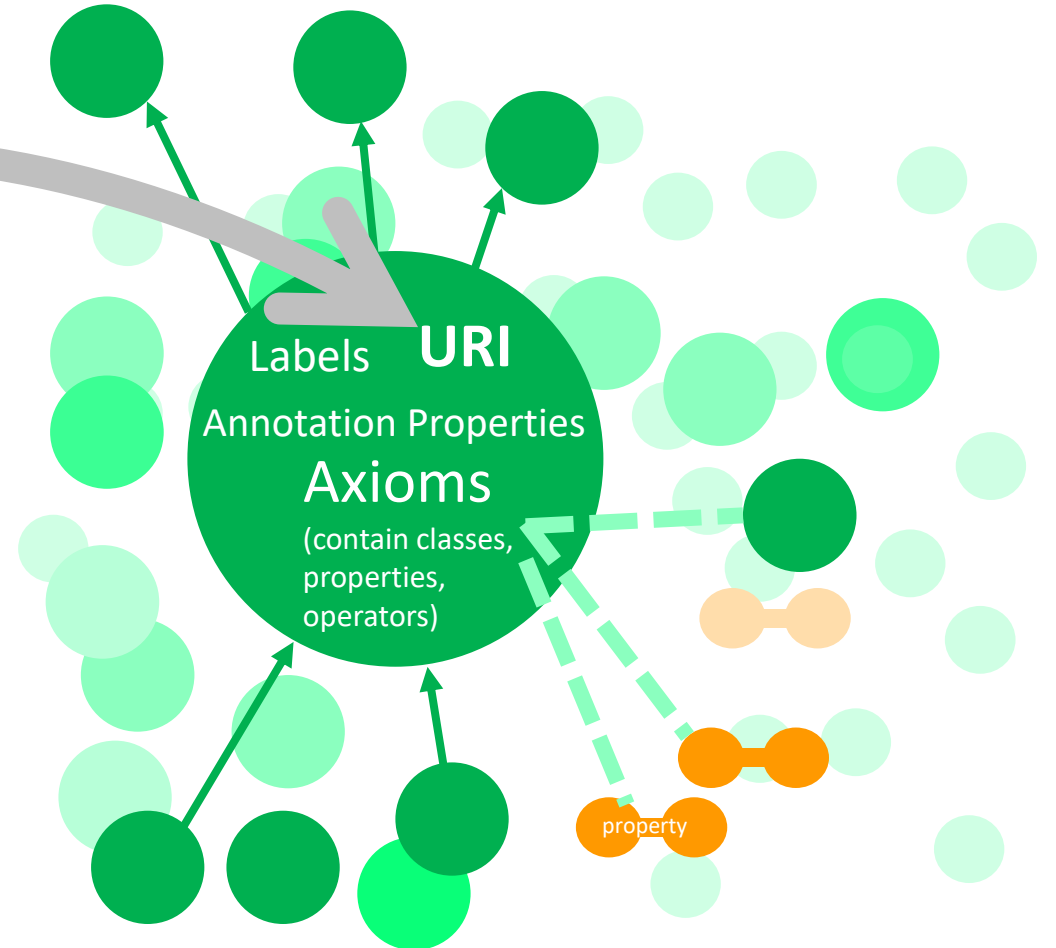
Selecting the ideal target ontologies

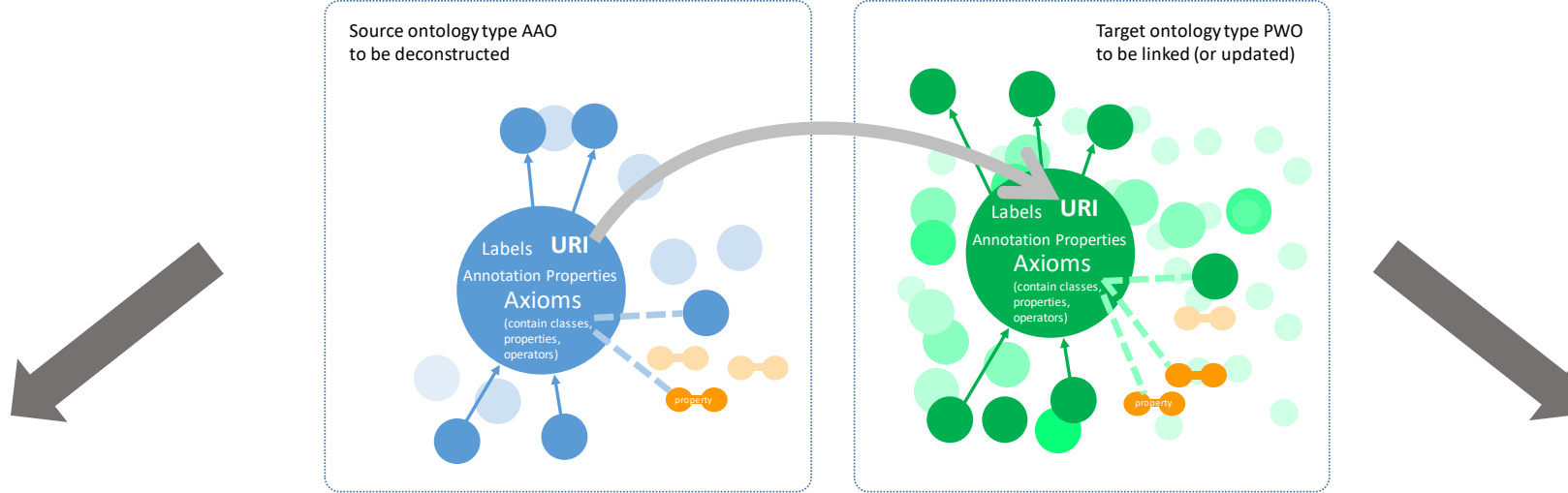
- Popular (= frequently used in other ontologies, applications and annotations)
- Well-maintained (community support)
- Created under a popular upper-level ontology
- Exclusion criteria
 - Artefacts that are not ontologies in a strict sense (not using OWL syntax and semantics)
 - Ontologies to which 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-Breis⁵ and Stefan Schulz¹

*Correspondence:

jose.vera-ramos@medunigraz.at

¹Institute 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

GRO recycling

GRO classes with suggested equivalences				
Class from existing GRO / New term	Candidate ontologies 1	Candidate ontologies 2	Resolution	Synonyms
phenotype	EFO	SNQ	Use class in EFO	
locus			Use class in OGI	locus
conformation	OGI			
function	BAO	BNAQ	Use class in BAO	
position	PATO	SWEEL	Use class in PATO	
genetic information		SWEET	Reconstruct	
feature			Use class in GO (GO_0097311)	
process	GO	MEQ	Use class in GO (GO_0097311)	

A	B	C	D	E
http://www.eurobiolab.org/ontology/GRO/ChromatinCofactor	http://purl.obolibrary.org/obo/CL_0000255			
http://www.eurobiolab.org/ontology/GRO/ChromatinComplex	http://amigo.geneontology.org/amigo/term/GO_0032993			
http://www.eurobiolab.org/ontology/GRO/ComplexOfProtein	http://amigo.geneontology.org/amigo/term/GO_1999504			
http://www.eurobiolab.org/ontology/GRO/NAIAssociatedSilencing	http://purl.obolibrary.org/obo/GO_0016442			
http://www.eurobiolab.org/ontology/GRO/ProteinComplex	http://amigo.geneontology.org/amigo/term/GO_0005681			
http://www.eurobiolab.org/ontology/GRO/RegulatoryRegion	http://semanticcience.org/resource/SIO_0000992			
http://www.eurobiolab.org/ontology/GRO/NucleicAcid	http://purl.obolibrary.org/obo/CHEBI_33695			
http://www.eurobiolab.org/ontology/GRO/NAIMolecule	http://purl.comics.ccrcc.uga.edu/ontologies/GlycoRNA_molecule			
http://www.eurobiolab.org/ontology/GRO/ChromosomalDNA	http://purl.obolibrary.org/obo/SO_0000954			
http://www.eurobiolab.org/ontology/GRO/Chromatin	http://amigo.geneontology.org/amigo/term/GO_0000795			
http://www.eurobiolab.org/ontology/GRO/Chromatin	http://amigo.geneontology.org/amigo/term/GO_0000791			
http://www.eurobiolab.org/ontology/GRO/heterochromatin	http://amigo.geneontology.org/amigo/term/GO_0000792			
http://www.eurobiolab.org/ontology/GRO/nucleosome	http://amigo.geneontology.org/amigo/term/GO_0000786			
http://www.eurobiolab.org/ontology/GRO/Chromatin	http://purl.obolibrary.org/obo/SO_0000140			
http://www.eurobiolab.org/ontology/GRO/Chromatin	http://purl.obolibrary.org/obo/SO_0000907			
http://www.eurobiolab.org/ontology/GRO/NABindingSite	http://purl.obolibrary.org/obo/SO_0001429			
http://www.eurobiolab.org/ontology/GRO/ProteinCodingGene	http://semanticcience.org/resource/SIO_0000985			
http://www.eurobiolab.org/ontology/GRO/NACodingGene	http://semanticcience.org/resource/SIO_0000986			

1

GRO classes with suggested equivalences

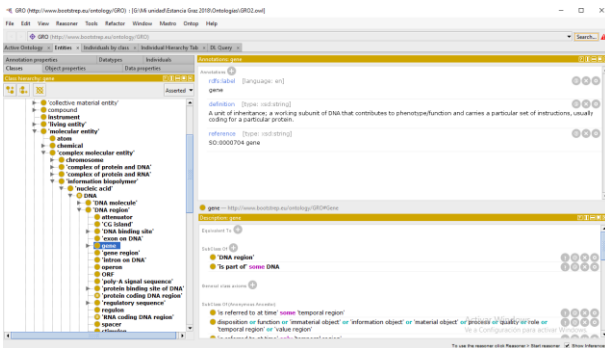
OLD_URI, NEW_URI

2

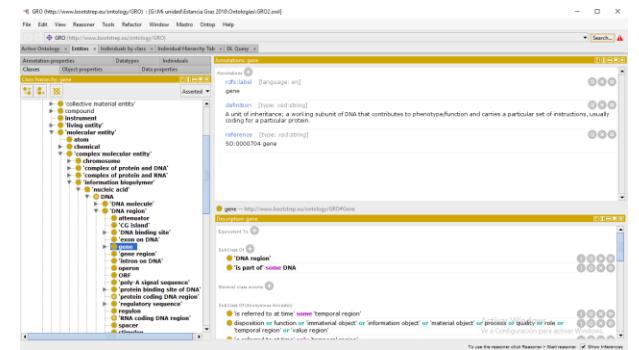
3



2



GRO



Recycled GRO

GRO recycling

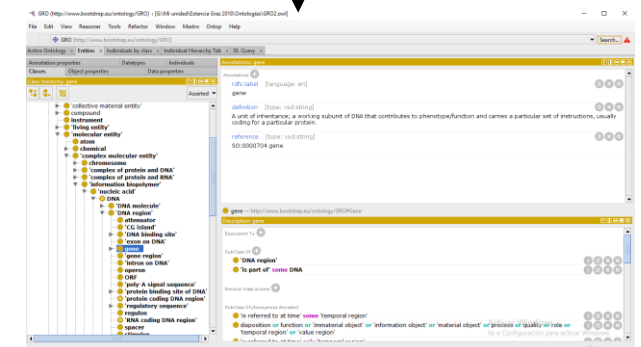
	A	B	C	D	E
21	http://www.ebi.ac.uk/ontology/GRO/EukaryoticCell	http://purl.obolibrary.org/obo/CL_000255			
22	http://www.ebi.ac.uk/ontology/GRO/ComplexOfProteinAn	http://amigo.geneontology.org/amigo/term/GO_0032993			
23	http://www.ebi.ac.uk/ontology/GRO/ComplexOfProteinAn	http://amigo.geneontology.org/amigo/term/GO_1950504			
24	http://www.ebi.ac.uk/ontology/GRO/RNAInterfering	http://purl.obolibrary.org/obo/GO_0016442			
25	http://www.ebi.ac.uk/ontology/GRO/Silicosis	http://amigo.geneontology.org/amigo/term/GO_0005681			
26	http://www.ebi.ac.uk/ontology/GRO/InformalOntology	http://semanticcience.org/resource/SIO_000092			
27	http://www.ebi.ac.uk/ontology/GRO/NucleicAcid	http://purl.obolibrary.org/obo/CHEBI_33595			
28	http://www.ebi.ac.uk/ontology/GRO/NAMolecule	http://glycomics.ccr.cba.ucla.edu/ontologies/GlycoCNA_molecule			
29	http://www.ebi.ac.uk/ontology/GRO/ChromosomalDNA	http://purl.obolibrary.org/obo/SC_0000594			
30	http://www.ebi.ac.uk/ontology/GRO/Chromatin	http://amigo.geneontology.org/amigo/term/GO_0007265			
31	http://www.ebi.ac.uk/ontology/GRO/Euchromatin	http://amigo.geneontology.org/amigo/term/GO_0000751			
32	http://www.ebi.ac.uk/ontology/GRO/Heterochromatin	http://amigo.geneontology.org/amigo/term/GO_0000752			
33	http://www.ebi.ac.uk/ontology/GRO/Nucleosome	http://amigo.geneontology.org/amigo/term/GO_0000756			
34	http://www.ebi.ac.uk/ontology/GRO/Heteromultimer	http://purl.obolibrary.org/obo/SC_0000149			
35	http://www.ebi.ac.uk/ontology/GRO/Cytoplasm	http://purl.obolibrary.org/obo/SC_0000307			
36	http://www.ebi.ac.uk/ontology/GRO/RNABindingSite	http://purl.obolibrary.org/obo/SC_0001429			
37	http://www.ebi.ac.uk/ontology/GRO/ProteinCodingGene	http://semanticcience.org/resource/SIO_000985			
38	http://www.ebi.ac.uk/ontology/GRO/RNACodingGene	http://semanticcience.org/resource/SIO_000986			

OLD_URI, NEW_URI

New_URI



label, definition, axioms



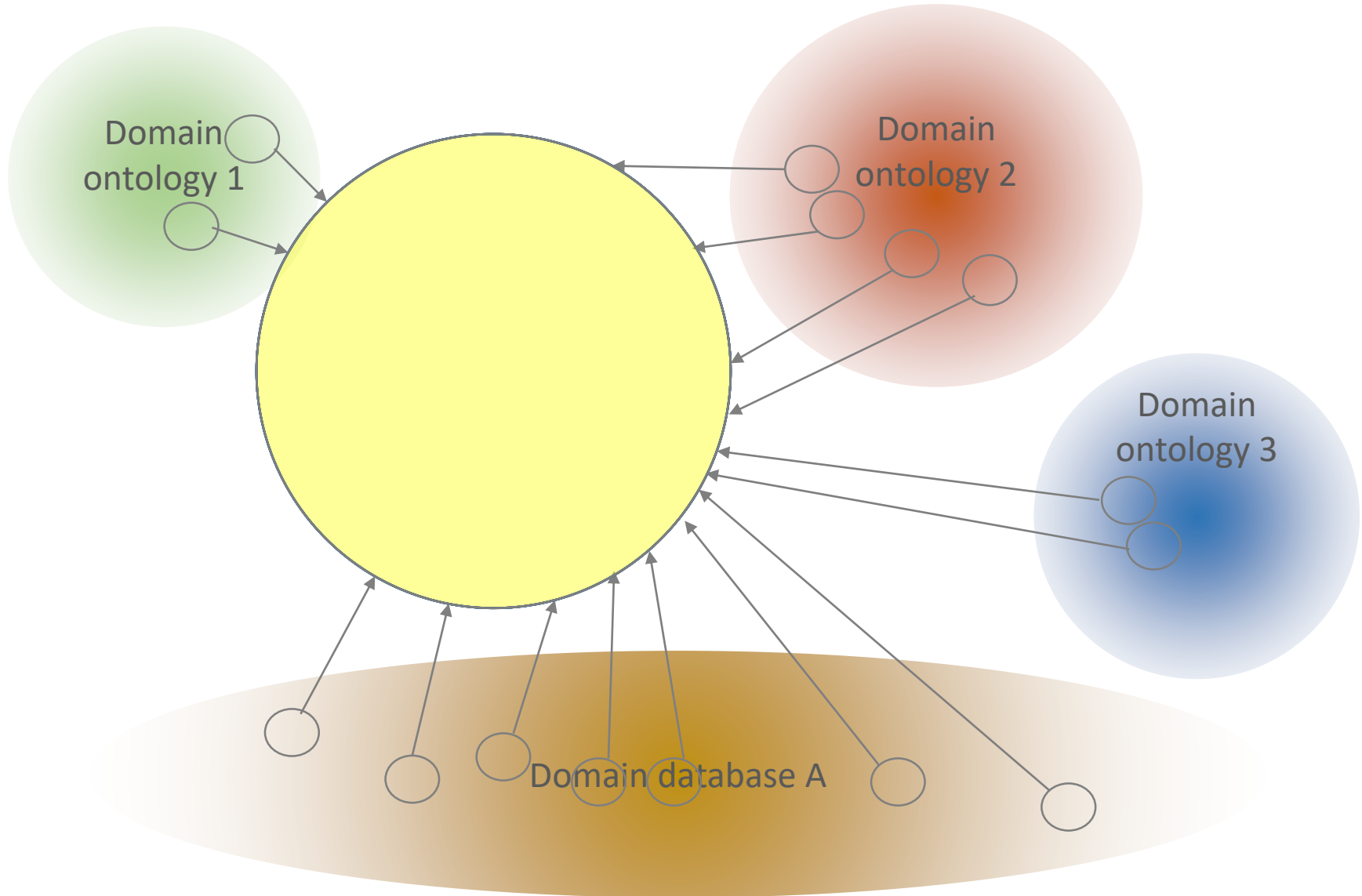
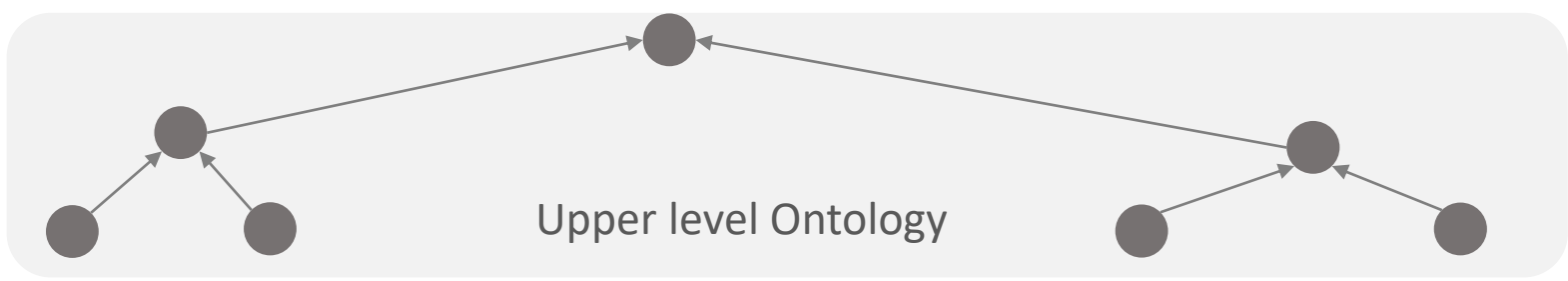
Recycled
GRO

¹<https://rdflib.github.io/sparqlwrapper/>

²Lamy 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)

