

Progress Towards an Ontology Mapping Service

**Ian Harrow¹, Ernesto Jimenez-Ruiz², Thomas Liener³, Simon Jupp³
and The Ontologies Mapping Project¹**

¹Ian Harrow Consulting, ²The Alan Turing Institute, London UK and ³EMBL-EBI, Hinxton, UK

SWAT4LS on 5th Dec 2018 in Antwerp

1. Pistoia Alliance

2. Application of Ontologies and Mappings

3. Progress Towards an Service

About Pistoia Alliance

Mission

- A global not-for-profit alliance of life science companies, vendors, publishers and academic groups
 - Lowering the barriers to innovation in Life Sciences R & D
 - www.pistoiaalliance.org for more information

Business value

- Precompetitive research brings value through...
 - Building new standards, tools and services
 - Sharing best practice with industry peers
 - Evaluation of tools and services
 - And much more....

Project portfolio pipeline

Active Portfolio

- Ontologies Mapping 
- Chemical Safety Library 
- Macromolecule Notation  HELM
Hierarchical Editing Language for Macromolecules
- Controlled Substance Compliance  CSCS
CONTROLLED SUBSTANCE COMPLIANCE SERVICE
- User Experience in Life Sciences (UXLS)  PISTOIA ALLIANCE
USER EXPERIENCE
FOR LIFE SCIENCES
- Antibody 3D structures in the PDB  AbVance
- Unified chemistry data model 

Developing Portfolio

- Faster Companion Diagnostics in the Regulatory Domain
- Blockchain
- AI Centre of Excellence for Life Science R&D
- Laboratory of the Future
- FAIR data management

Outline

1. Pistoia Alliance

2. Application of Ontologies and Mappings

3. Progress Towards an  Service

Ontology Features: Example from Gene Ontology

Class: (+)-2-epi-prezizaene synthase activity

1) Class Terms (Controlled Vocabulary or TBox)

Term IRI: http://purl.obolibrary.org/obo/GO_0102201

Definition: Catalysis of the reaction: 2-cis,6-trans-farnesyl diphosphate <=> (+)-2-epi-prezizaene + diphosphoric acid

Annotations

- database_cross_reference:MetaCyc:RXN-12117(= ontology mapping)
- has_obo_namespace:molecular_function
- id:GO:0102201

3) Identifier

Class Hierarchy

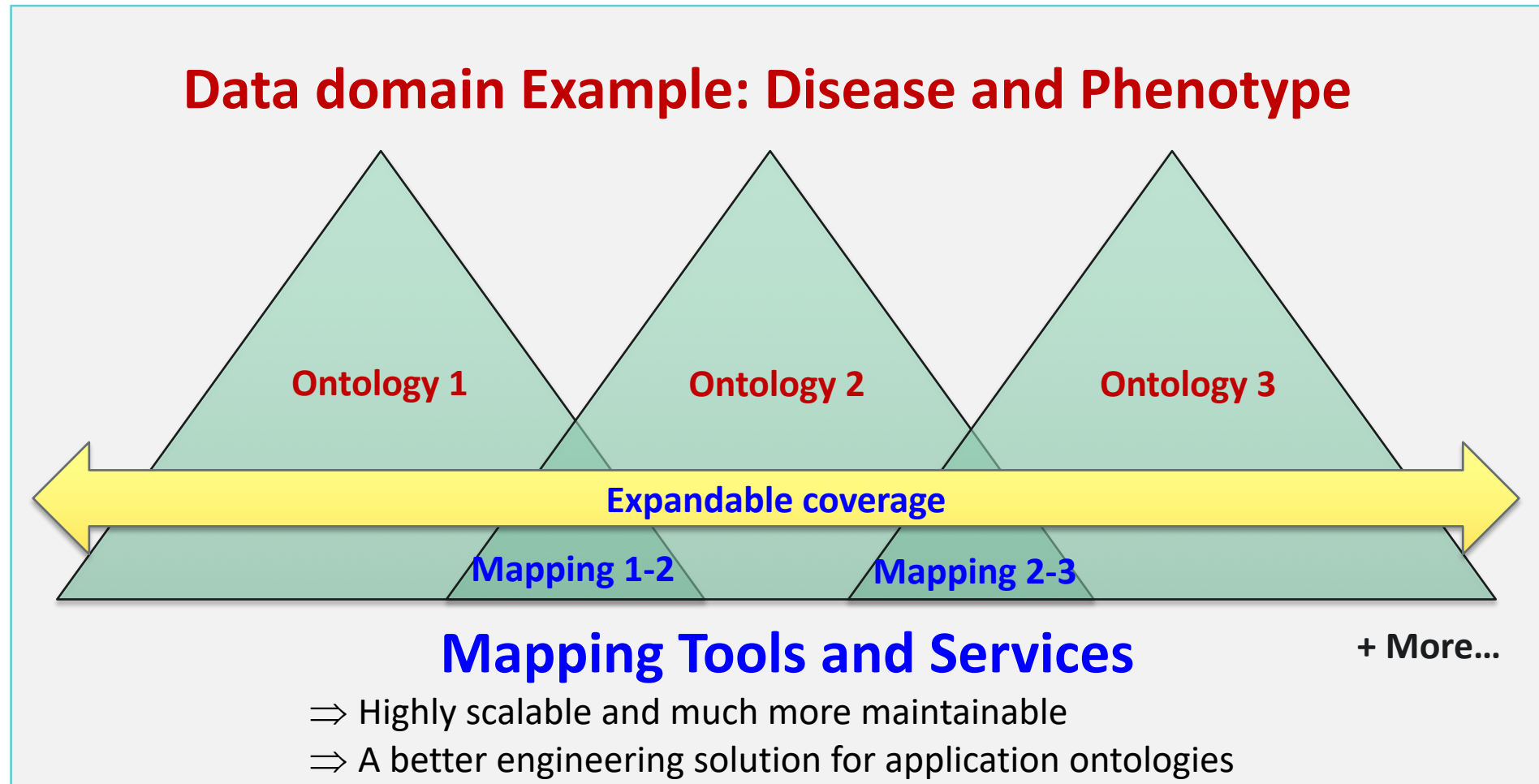
Thing
+ [molecular_function](#)
+ [catalytic activity](#)
+ [lyase activity](#)
+ [carbon-oxygen lyase activity](#)
+ [carbon-oxygen lyase activity, acting on phosphates](#)
- [3-dehydroquinase synthase activity](#)
- [6-pyruvoyltetrahydropterin synthase activity](#)
- [chorismate synthase activity](#)
- [threonine synthase activity](#)
- [methylglyoxal synthase activity](#)
+ [terpene synthase activity](#)
- [germacradienol synthase activity](#)
- [S-linalool synthase activity](#)
- [R-linalool synthase activity](#)
- [ent-cassa-12,15-diene synthase activity](#)
- [stemar-13-ene synthase activity](#)
- [syn-pimara-7,15-diene synthase activity](#)
[more...](#)
- [\(+\)-2-epi-prezizaene synthase activity](#)

2) Class Hierarchy
(Structure or ABox)





Ontology:
1) Class Terms
2) Class Hierarchy
3) Identifier

Source: <http://www.ontobee.org>

What is Ontologies Mapping?




Application of ontologies and mappings

- Pharma executives now consider data as a valuable corporate assets to enable digital transformation
 - Data integration throughout an enterprise e.g. 
 - Horizontal Terminology Services
- Data Technology companies bridge the gap between “big data” and “innovative biological insight”
 - Data curation, valuation and governance
 - E.g. Eaglecore knowledge management platform at 
- Semantic analytics companies harness unstructured data
 - Data extraction and building knowledge with text mining
 - E.g.  and  platforms

Ontologies and mappings bring structure to linked data stores

Outline

1. Pistoia Alliance
2. Application of Ontologies and Mappings
3. Progress Towards an  Service

3. The Ontologies Mapping Challenge

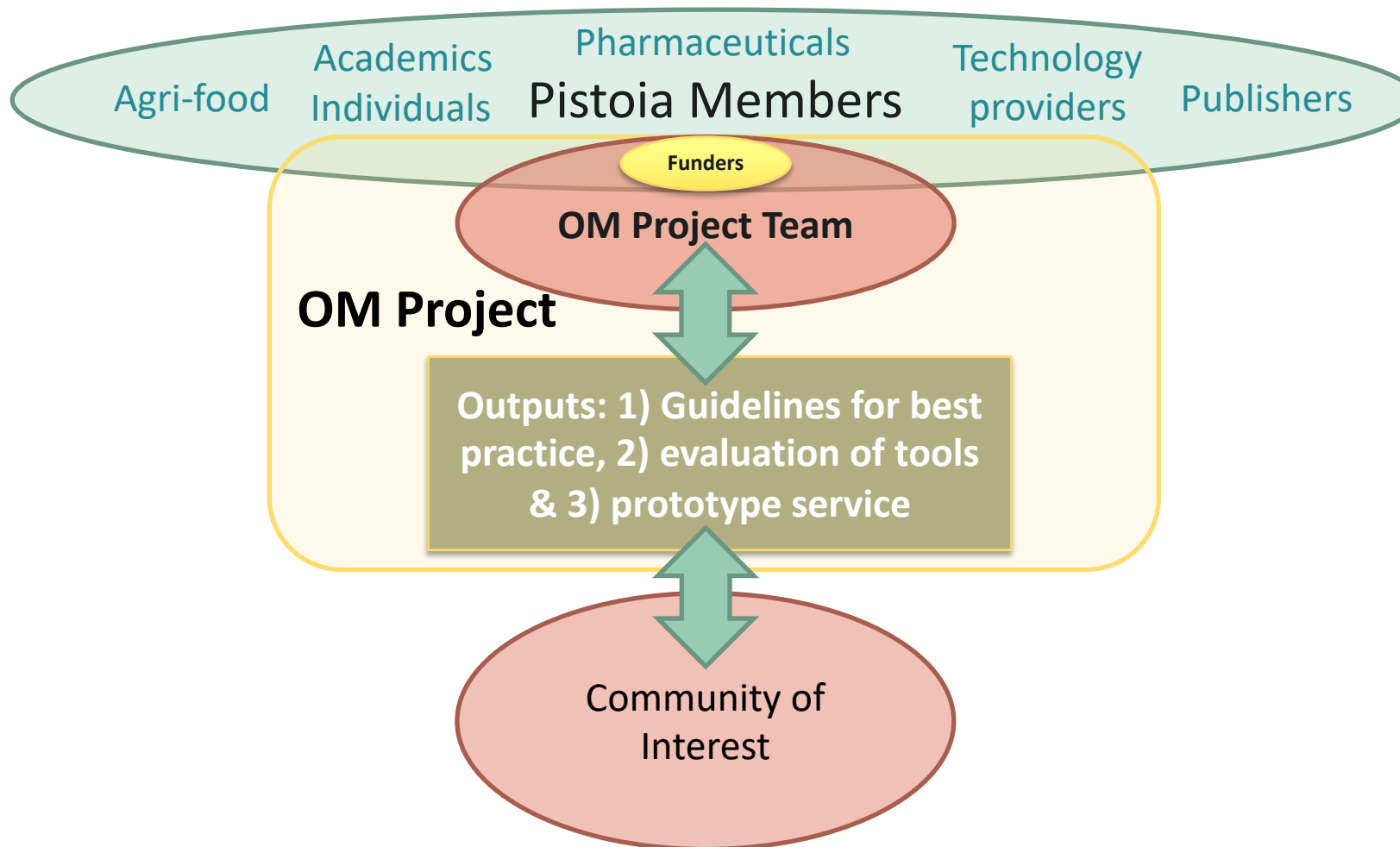
- Ontologies can include hierarchical relationships; taxonomies; classifications and vocabularies
- They underpin numerous applications such as semantic search, linked data integration and text mining
- Ontologies are the “**smart glue**” of linked data to semantically enable knowledge management

BUT...

- Ontologies and their mappings are very costly to curate
- Many varying ontologies overlap in the **same data domain** e.g. disease and phenotype
- Need better practice, tools and services to manage and apply ontologies, including **how they map to each other**

 **The Ontologies Mapping Project**

Ontologies Mapping Project Overview

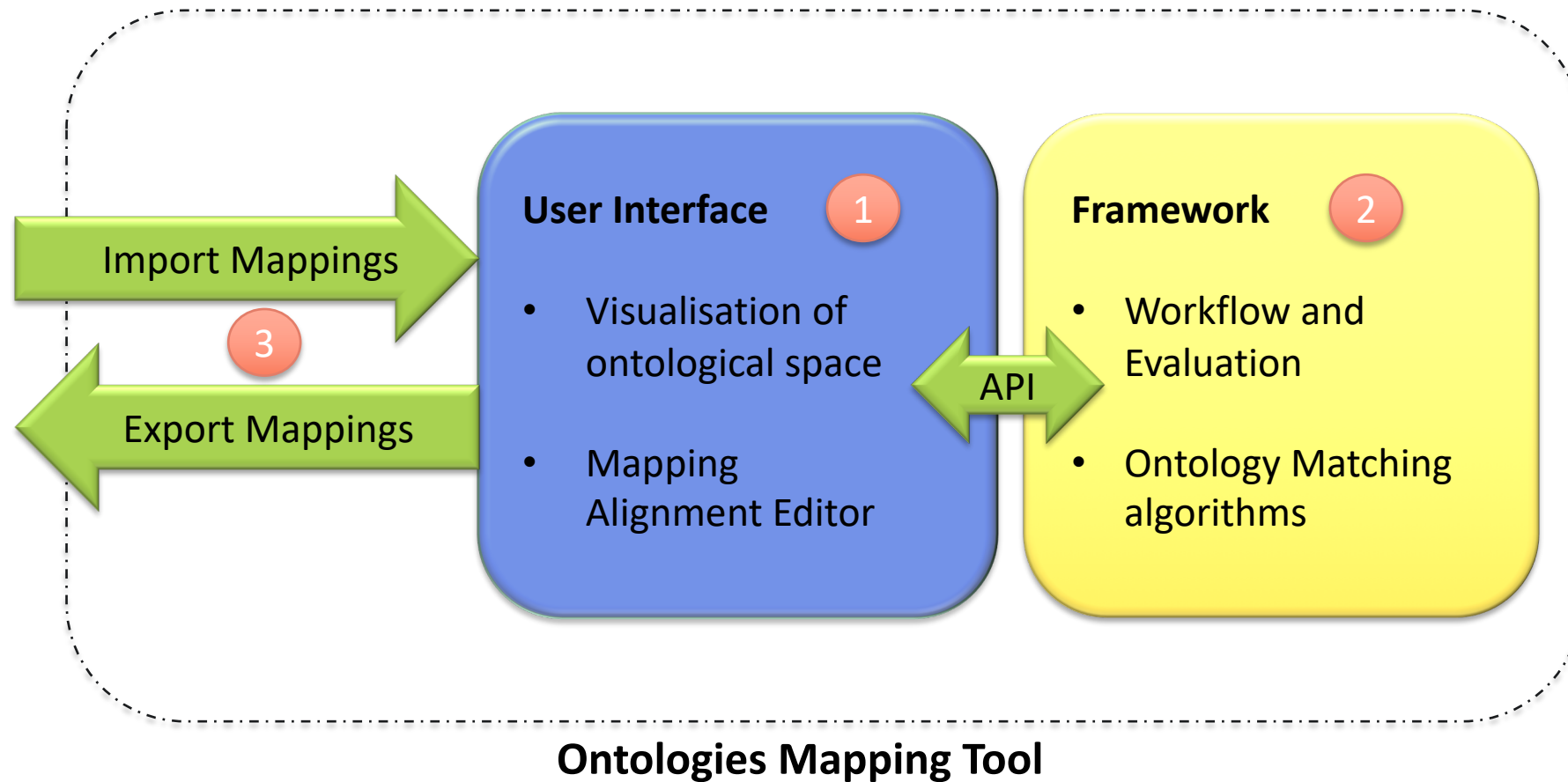


Guidelines for Selection of Ontologies

Ontology: Guideline:	Human Disease ontology (DOID)	Human Phenotype (HPO)	Mammalian Phenotype (MP)	Clinical Terms SNOMED-CT
1. Format (Common)	Okay	Okay	Okay	Okay
2. URI (Identifier space)	Okay	Okay	Okay	Okay
3. Versioning	Okay	Okay	Okay	Okay
4. Documentation	Okay	Okay	None found	Okay
5. Users (Documented)	Okay	Okay	Okay	Okay
6. Authority (Locus of)	Okay	Okay	Okay	Okay
7. Maintenance	Okay	Okay	Okay	Okay
8. License (Open)	Okay	Okay	Okay	Restrictions
Seven more..... (Total = 15)	Okay	Okay	Okay	Okay (mostly!)

- These guidelines are accessible from a public wiki:-
<https://pistoiaalliance.atlassian.net/wiki/display/PUB/Ontologies+Mapping+Resources>
- They align with the principles found at the OBO Foundry:- <http://www.obofoundry.org>

Ontologies Mapping Tool Overview



Detailed requirements are available on the OM project public wiki:-

<https://pistoiaalliance.atlassian.net/wiki/display/PUB/Ontologies+Mapping+Resources>

Tool Requirements & Evaluation of Capability

Functional Requirements	Academic 1	Academic 2	Academic 3	Commercial 2	Commercial 1	Academic 4	Commercial 3
1.1.1. Numerous view options	1	1	1	1	2	1	2
1.2.1. Improving Alignments	0	0	1	1	2	2	2
1.2.2. Matching correspondence	0	1	1	2	0	2	2
1.2.3. Edit mapping suggestions	0	0	2	2	2	2	2
1.2.4. Tracking of modifications	0	1	0	1	2	1	2
1.2.5. Definition of context	0	0	0	1	0	0	0
2.1.1. Workflow	1	0	1	1	2	2	2
2.1.2. Evaluation metrics	1	0	2	0	2	2	1
2.2.1. Supports extensibility	0	1	1	1	0	2	2
3.1.1. Import equivalence mappings	0	2	2	2	2	2	2
3.1.2. Import source ontologies	2	2	2	2	2	2	2
3.1.3. Use of external data sources	2	0	1	2	2	2	0
3.2.1. Export equivalence mappings	1	2	2	2	2	2	2
3.2.2. Mapping metadata & docs	1	1	0	0	2	2	1
None-Functional requirements							
1. No License restrictions for use	2	2	2	2	2	2	2
2. Current Availability & Maintenance	1	1	2	2	2	1	2
3. Standalone and web service	2	1	2	2	2	1	2
LogMap, AML, OLS/OXO, YAM++ Mondeca, Infotech, fluidOps	41%	44%	65%	71%	82%	82%	82%
Key:-	Yes = 2	Partial = 1	Expected = 0	No = 0			

Detailed requirements and results are available on the OM project public wiki:-

<https://pistoiaalliance.atlassian.net/wiki/display/PUB/Ontologies+Mapping+Resources>

Ontology Alignment Evaluation Initiative

<http://oaei.ontologymatching.org>



- 2016 Campaign – publication on phenotype track
 - “Matching disease and phenotype ontologies in the ontology alignment evaluation initiative” Harrow et al. Journal of Biomedical Semantics (2017) 8:55 <https://doi.org/10.1186/s13326-017-0162-9>
- Numerous tracks include one for Disease and Phenotype
 - Pistoia Alliance Ontologies Mapping project organise 2 mapping tasks:-
 - Human Phenotype (HP) Ontology vs. Mammalian Phenotype (MP) Ontology
 - Human Disease Ontology (DOID) vs. Orphanet & Rare Diseases Ontology (ORDO)
- Phenotype track repeated for 2017 and 2018 Campaigns
 - ~50% of ~20 participating systems completed either or both tasks
 - Consistent top performing systems: AML and LogMap

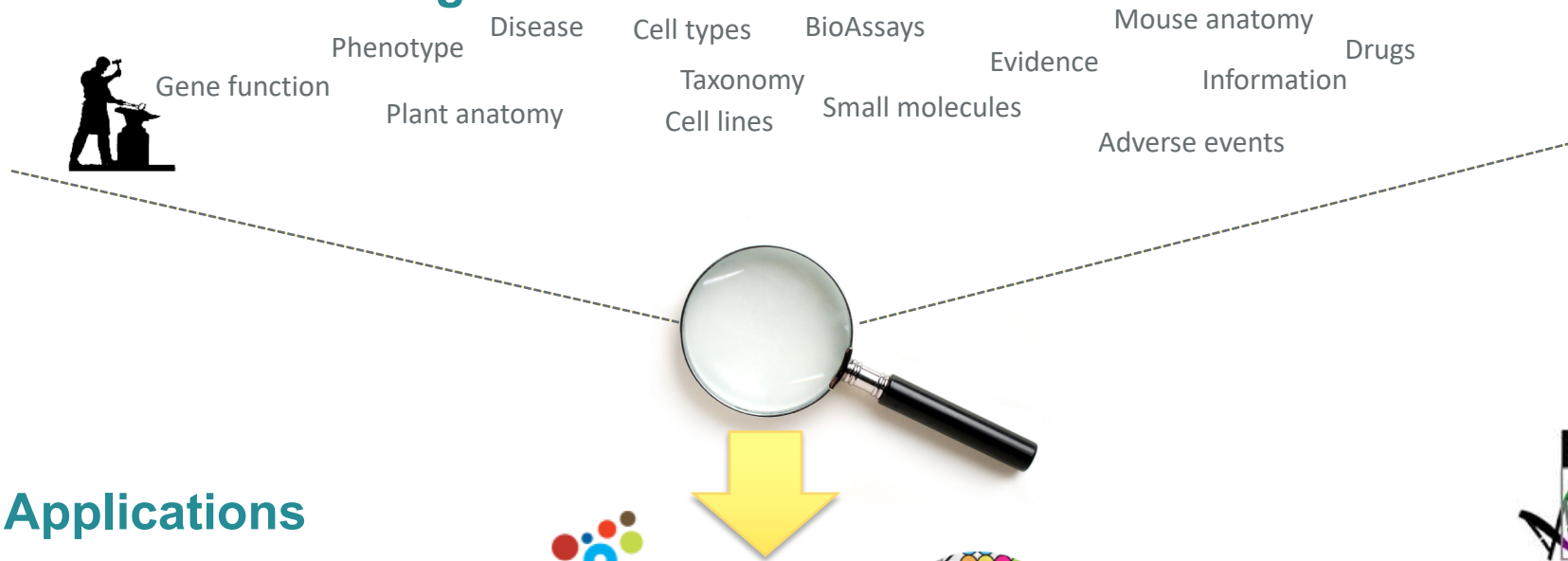
Top performers for OAEI 2017 in phenotype track

OM algorithm	Track Task	Total Equivalence Mappings	Precision Silver 3 Equiv mappings	Recall Silver 3 Equiv mappings	F-Score Silver 3 Equiv mappings	Sum F Scores Silver 3 Equiv mappings
AML	HP-MP	2029	0.822	0.951	0.882	3.791
AML	DOID-ORDO	4779	0.475	0.626	0.919	
AML	HP-MESH	5638	0.677	0.805	0.992	
AML	HP-OMIM	6681	0.624	0.768	0.998	
DiSMATCH AR	HP-MP	2378	0.500	0.678	0.576	3.144
DiSMATCH AR	DOID-ORDO	3130	0.539	0.603	0.684	
DiSMATCH AR	HP-MESH	9161	0.385	0.542	0.917	
DiSMATCH AR	HP-OMIM	7356	0.549	0.701	0.967	
DiSMATCH TR	HP-MP	2331	0.517	0.687	0.590	3.183
DiSMATCH TR	DOID-ORDO	3089	0.545	0.606	0.682	
DiSMATCH TR	HP-MESH	9138	0.389	0.547	0.924	
DiSMATCH TR	HP-OMIM	7680	0.537	0.696	0.988	
LogMap	HP-MP	2124	0.767	0.929	0.840	3.149
LogMap	DOID-ORDO	2396	0.903	0.890	0.876	
LogMap	HP-MESH	2291	0.869	0.649	0.518	
LogMap	HP-OMIM	7202	0.531	0.672	0.915	
LogMapBio	HP-MP	2204	0.749	0.941	0.834	3.291
LogMapBio	DOID-ORDO	2620	0.845	0.871	0.897	
LogMapBio	HP-MESH	2948	0.810	0.703	0.621	
LogMapBio	HP-OMIM	7725	0.508	0.659	0.939	
BioPortal LOOM	HP-MP	696	0.999	0.396	0.567	2.599
BioPortal LOOM	DOID-ORDO	1237	0.998	0.666	0.500	
BioPortal LOOM	HP-MESH	2466	0.994	0.776	0.637	
BioPortal LOOM	HP-OMIM	3768	0.992	0.941	0.895	

Source: This data is from the 3 vote consensus

Ontologies at EMBL-EBI

Biomedical ontologies







Applications



Source: Dr Simon Jupp (EMBL-EBI)

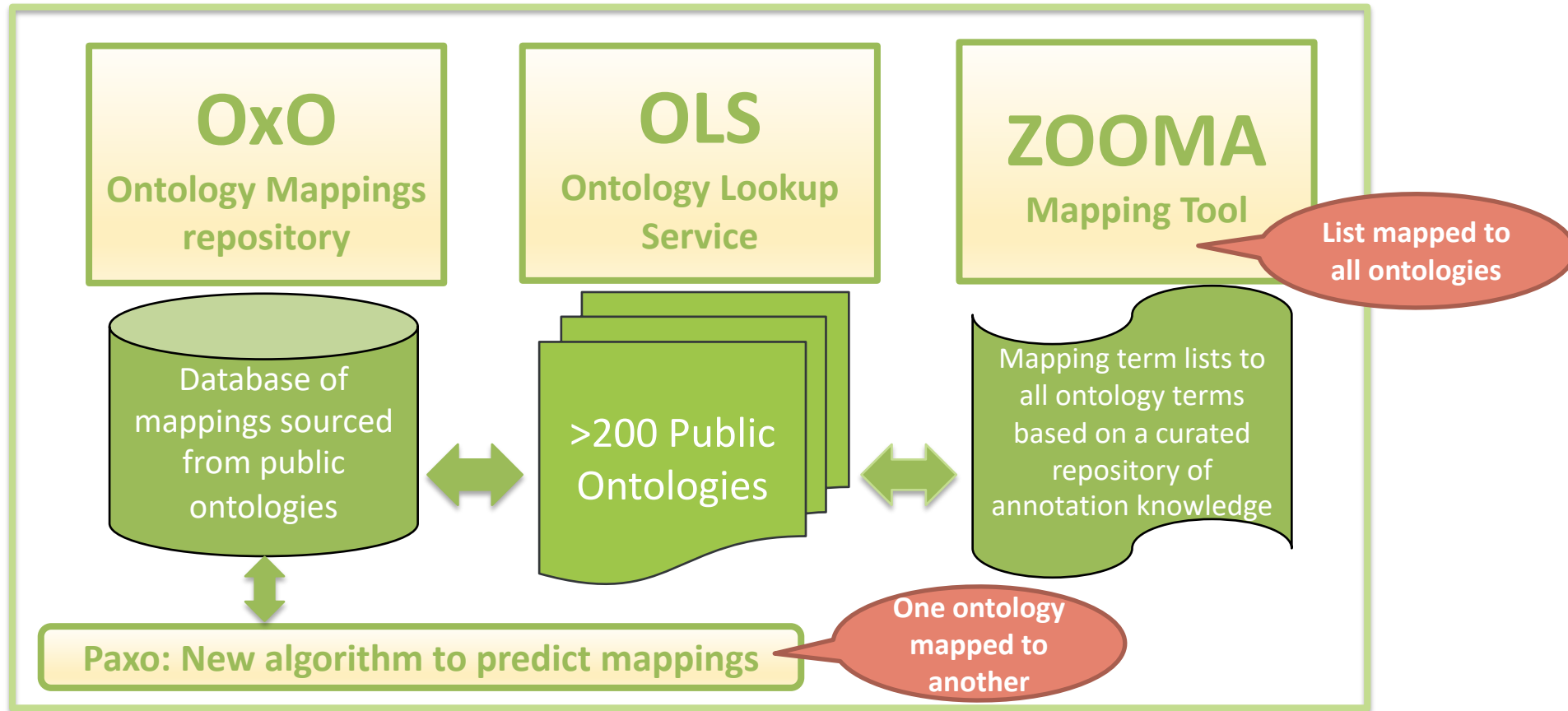
Building an Ontology Toolkit and Services

 <small>ONTOLOGY SEARCH</small> Ontology Lookup Service	Search/Visualise ontologies
 <small>ONTOLOGY ANNOTATION</small> Zooma	Annotate data
 <small>ONTOLOGY MAPPING</small> OxO	Ontology Mappings
 <small>ONTOLOGY CREATION</small> Webulous	Create new ontology content

Source: Dr Simon Jupp (EMBL-EBI)

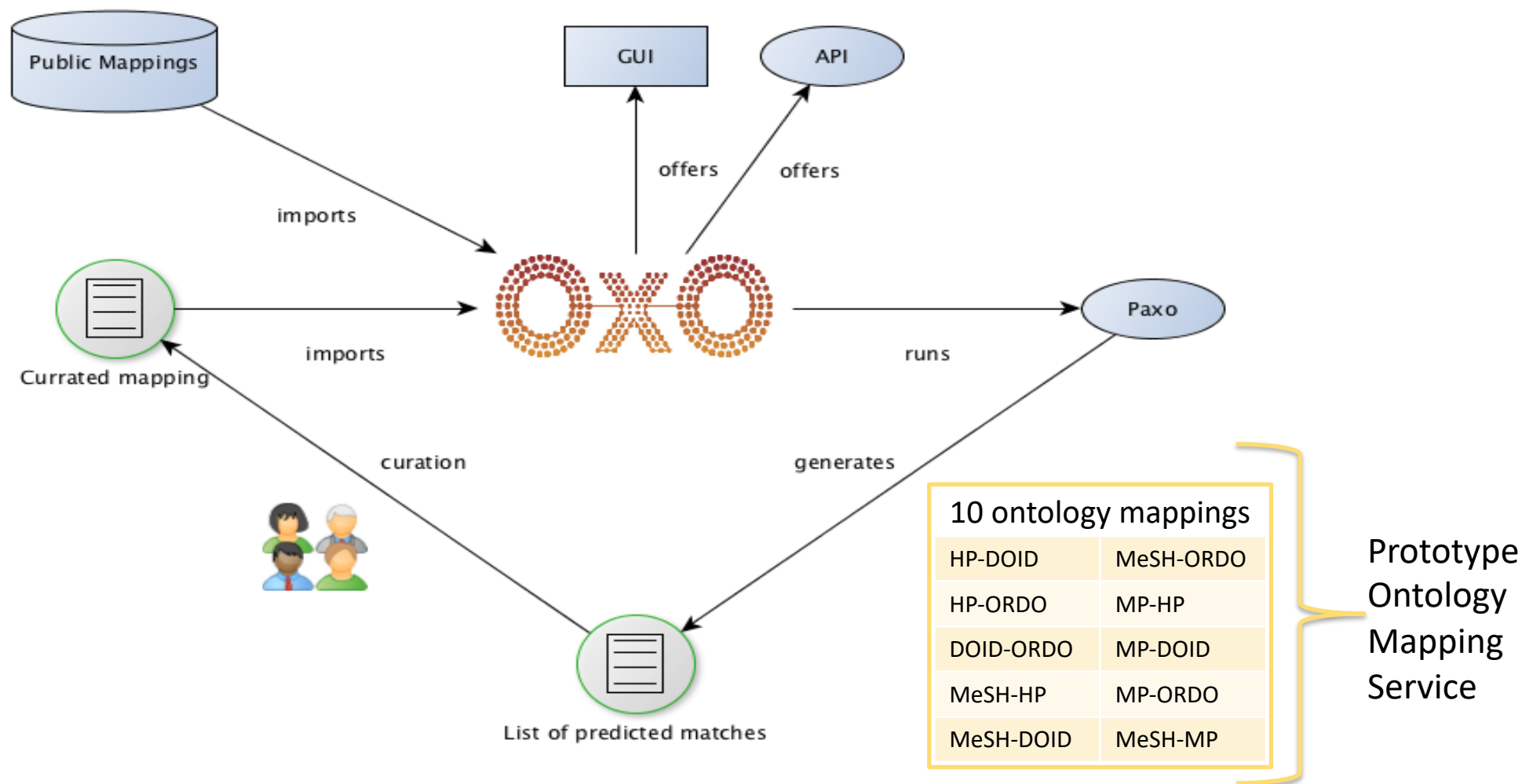
Grants: ELIXIR-EXCELERATE 676559 & CORBEL 654248

How the Ontology Toolkit & Services fit together



- Collaborate with EMBL-EBI to build on their existing services
- Develop a new algorithm to predict mappings between any two ontologies

Relationship between OxO mapping repository and the new mapping algorithm, Paxo



OxO Ontology Mappings Service



Home < Ontology Xref Service

https://www.ebi.ac.uk/ 67%

EMBL-EBI

Services Research Training About us

Ontology Xref Service

OxO Documentation About Contact Us

BETA This site is still under development

Welcome to the EMBL-EBI Ontology Xref Service (OxO).

OxO is a service for finding mappings (or cross-references) between terms from ontologies, vocabularies and coding standards. OxO imports mappings from a variety of sources including the [Ontology Lookup Service](#) and a subset of mappings provided by the [UMLS](#). We're still developing the service so please [get in touch](#) if you have any feedback.

Search OxO by term id

Choose a target (optional)

Datasources

Mapping Distance: 1 2 3

Enter list of identifiers: Show me some examples...

List of Identifiers Or labels

OxO summary

To use OxO either search for mappings using a particular identifier (e.g. MESH:D009202) or select a datasource below to view all mappings between datasources.

Jump to a datasource

Human

Human Developmental Stages

Human Disease Ontology

Human developmental anatomy, abstract

human phenotype ontology

Datasource < Ontology Xref Service

https://www.ebi.ac.uk/ 67%

EMBL-EBI

Services Research Training About us

Ontology Xref Service

OxO Documentation About Contact Us

BETA This site is still under development

Datasource

Human Disease Ontology

An ontology for describing the classification of human diseases organized by etiology.

Version info: Last updated in the ontology lookup service on 2017-03-24T01:03:56.177+0000

Licence info: <https://creativecommons.org/licenses/by/4.0/>

Prefix: [D001](#)

Type: ONTOLOGY View in OLS

Mappings

Mapping Distance: 1 2 3

Mappings by target

Export mapping

1,540 matches

Human Disease Ontology

Search < Ontology Xref Service

https://www.ebi.ac.uk/ 67%

EMBL-EBI

Services Research Training About us

Ontology Xref Service

OxO Documentation About Contact Us

BETA This site is still under development

Mapping results

Select a term to see more information. The evidence column tells you how many times we have seen this mapping and the distance is a how many hops across mappings you need to go to find this mapping. Distance 1 is a direct mapping, the greater the distance the less likely it is that a mapping holds true. Max distance is 3.

Showing 1 to 50 of 1,540 entries

Input	Mapped id	Id source	Evidence	Distance
D001:0050335 (bradyopsia)	Orphanet:75374 (Bradyopsia)	Orphanet	1	1
D001:0050428 (nonepidermolytic palmoplantar keratoderma)	Orphanet:2337 (Non-epidermolytic palmoplantar keratoderma)	Orphanet	2	1
D001:0050430 (multiple endocrine neoplasia type 2A)	Orphanet:247698 (Multiple endocrine neoplasia type 2A)	Orphanet	2	1
D001:0050431 (arrhythmogenic right ventricular cardiomyopathy)	Orphanet:247 (Arrhythmogenic right ventricular cardiomyopathy)	Orphanet	2	1
D001:0050432 (arrhythmogenic right ventricular cardiomyopathy)	Orphanet:217636 (Familial isolated arrhythmogenic right ventricular dysplasia)	Orphanet	2	1
D001:0050433 (Asperger syndrome)	Orphanet:1162	Orphanet	2	1
D001:0050434 (Andersen-Tawil syndrome)	Orphanet:37553 (Cardiarrhythmic periodic paralysis)	Orphanet	2	1
D001:0050435 (Usher syndrome)	Orphanet:886 (Usher syndrome)	Orphanet	2	1
D001:0050440 (familial partial lipodystrophy)	Orphanet:98306 (Familial partial lipodystrophy)	Orphanet	1	1
D001:0050444 (hereditary mucosal leukokeratosis)	Orphanet:171723 (White sponge nevus)	Orphanet	2	1
D001:0050453 (lissencephaly)	Orphanet:102009 (Classic lissencephaly)	Orphanet	2	1
D001:0050454 (periventricular nodular heterotopia)	Orphanet:96892 (Periventricular nodular heterotopia)	Orphanet	2	1
D001:0050455 (Newey-Dicks syndrome)	Orphanet:96892 (Periventricular nodular heterotopia)	Orphanet	2	1

Download CSV TSV JSON

Orphanet

Optimisation of the predicted ontology mappings for unique matches: evaluation of quality

Mapping	Predicted matches total	Predicted in silver	Silver standard	Missed matches	Recall	Additional matches	Precision for additions (N=60)
mesh_mp	796	280	282	2	99.29%	516	96.70%
mesh_doid	2173	1253	1265	12	99.05%	920	86.70%
mesh_hp	1400	724	734	10	98.64%	676	60.00%
ordo_mesh	970	632	664	32	95.18%	338	85.00%
hp_doid	1976	1104	1348	244	81.90%	872	33.30%
ordo_doid	2732	2044	2553	509	80.06%	688	66.70%
ordo_hp	1305	593	752	159	78.86%	712	63.30%
ordo_mp	550	138	185	47	74.59%	412	73.30%
mp_doid	1087	310	465	155	66.67%	777	33.00%
mp_hp	2600	1318	2185	867	60.32%	1282	53.30%

Silver standard is a consensus from a panel of ontology mapping algorithms from OAEI2017

Optimisation of the predicted ontology mappings for multiple matches: evaluation of quality

Mapping	Predicted matches total	Predicted in silver	Silver standard	Missed matches	Recall	Additional matches	Precision for additions (N=60)
mesh_mp	868	282	282	0	100.00%	586	88.30%
mesh_doid	2280	1261	1265	4	99.68%	1019	81.70%
mesh_hp	1519	728	734	6	99.18%	791	60.00%
hp_doid	3636	1156	1348	192	85.76%	2480	40.00%
ordo_doid	5375	2178	2553	375	85.31%	3197	73.30%
ordo_hp	2382	623	767	144	81.23%	1759	68.30%
ordo_mesh	1079	632	782	150	80.82%	447	88.30%
ordo_mp	958	141	185	44	76.22%	817	56.70%
mp_doid	2473	320	465	145	68.82%	2153	30.00%
mp_hp	7190	1446	2128	682	67.95%	5744	43.30%

Silver standard is a consensus from a panel of ontology mapping algorithms from OAEI2017

Summary

- Guidelines to select ontologies prior to application
- Requirements vs. capabilities of commercial and academic mapping tools
- Phenotype track for OAEI challenge to determine top performing algorithms
- Progress towards a prototype Ontology Mapping Service

Significance

- Predicted Ontology Mappings from algorithms and validated by manual curation gives extended their coverage across a domain to support semantically-enabled applications

Plan for Next Phase

- Build Ontology Mappings to support biology and chemistry lab data analytics

Acknowledgements



Sponsors (Phase)

- BIOVIA 3DS (2,3)
- GSK (1,2,3)
- Roche/Genentech (1,2,3)
- Amgen (3)
- AZ/MedImmune (3)
- Accenture (3)
- Bayer (3)
- Merck & Co (1,2)
- Novartis (1,2)

Project team

- Ian Harrow (Project Manager)
- Martin Romacker (Roche)
- Rama Balakrishnan (Genentech)
- Andrea Splendiani (Novartis)
- Peter Woollard/Simon Thornber (GSK)
- Scott Markel (BIOVIA)
- Siddhartha Mehta (Accenture)
- Chris Piddington/David Landry (Amgen)
- Tom Plasterer (AstraZeneca)
- Mathew Woodward (Medimmune)
- Stefan Negru/Jindrich Mynarz (Merck & Co)
- Thomas Liener/Simon Jupp (EMBL-EBI)
- Jabe Wilson (Elsevier)
- Berenice Wulbrecht (Ontoforce)
- Christian Senger/Martin Koch (Osthus)
- Heiner Oberkamp (Osthus)
- Yasmin Alam-Faruque (Eagle Genomics)
- Rolf Grigat/Erfan Younesi (Bayer)
- CJ Farham Hameed (Pfizer)
- Rainer Winnenbourg (Abbvie)
- Hans Garritzen (MediSapiens)
- Jane Reed/Andrew Winter (Linguamatics)
- Jane Lomax/James Malone (SciBite)
- Carmen Nitsche/Nick Lynch (Pistoia)