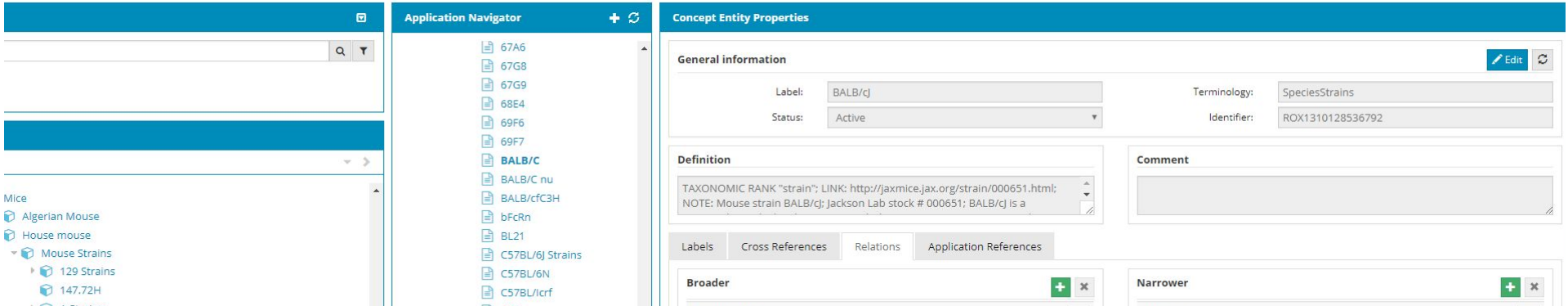
Semantic Terminology Management for Applications: Contextualized SKOS-XL

Andreas Thalhammer, Martin Romacker, and Joachim Rupp

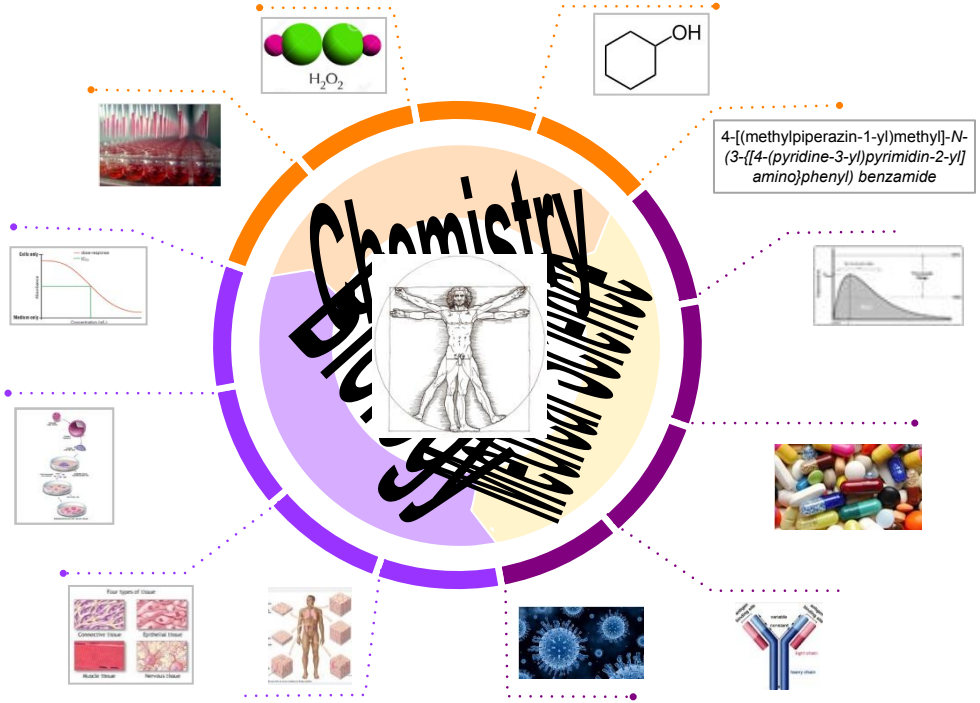
Pharma Research and Early Development Informatics
Solution Delivery & Architecture, Roche Innovation Center Basel

16th International Semantic Web Conference, Vienna, Austria 23.10.2017



Life Sciences Research and Development

Knowledge Space and Information Objects in R&D



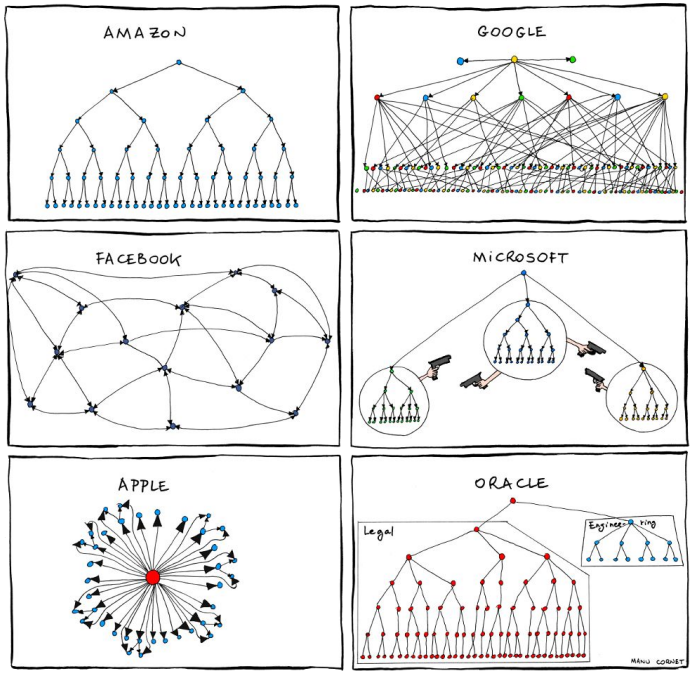
Huge universe of scientific and business objects to be represented

Slide courtesy of Martin Romacker

Main Challenge

Organization-wide terminology coverage is hard to achieve

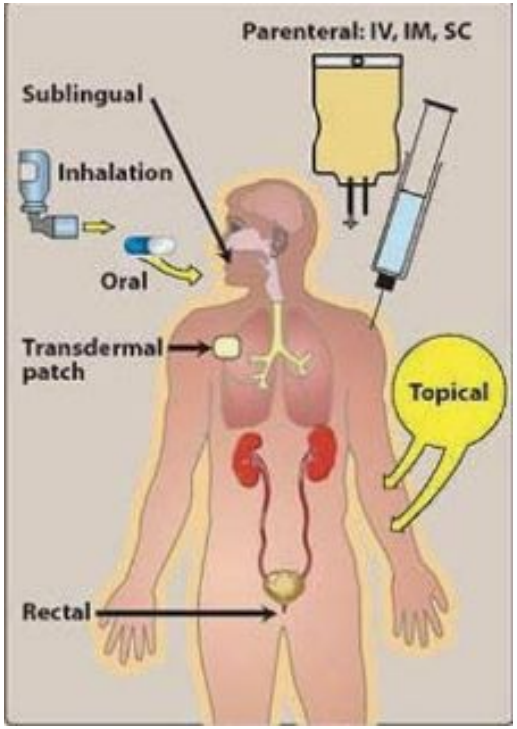
- We maintain hundreds of different applications used to capture data.
- Applications and additional data can come from different (internal and external) solutions.
- Application landscape reflects organizational structure (Conway's law).



Source: http://bonkersworld.net/organizational-charts

Terminology Integration

Basic Level of Semantic Data Integration



LIPA_DISP_VALUE
ROUTE_IV_INFUSION
ROUTE_IP
PO
SC
INTRANALSAL
in vitro
Intravenous
Intramuscular
Intracisternal
Intraocular
Intraperitoneal
Topical
Food addivitive, peroral
Intracerebroventricular
Inhaled aerosol or gas
Intranasal
Dietary
Intratracheal
Oral by mouth (per orum)
Intravaginal
Subcutaneous
Transdermal
UNKNOWN
ROUTE_PO_ADMIX
ROUTE_PO_RAPID
ROUTE_TOPICAL
ROUTE_IVT
TOPICAL
ORTHER
intravitreal

LIPA_DISP_VALUE
GAVAGE
INFUSION
INTUBATION
UNKNOWN
ROUTE_PO_ADMIX
ROUTE_PO_RAPID
RAPID
WATER ADMIX
ADMIX
NO VALUE
DIETARY
Eye Drop
BOLUS
INTRANALSAL
ROUTE_SC
ROUTE_PO_GAVAGE
TP
MINIPUMP
ROUTE_IP
SC
FOOD ADMIX
PO
IP
ROUTE_IV_INFUSION
CAPSULE

Slide courtesy of Martin Romacker

Data Quality

What are we talking about?

- Data quality starts at application design.
- Freetext fields are the worst possible solution.
- Favorable elements are:
  drop down fields, radio buttons, checkboxes...

....

Semantic Thesaurus Management Tools

- Typical vendor solution:

"Go to our SPARQL endpoint and execute

```
SELECT * WHERE {
    ?disease rdf:type skos:Concept .
    ?disease skos:inScheme roche:Indication .
    ?disease skos:prefLabel ?label .
}
```

You will get all diseases of the thesaurus."

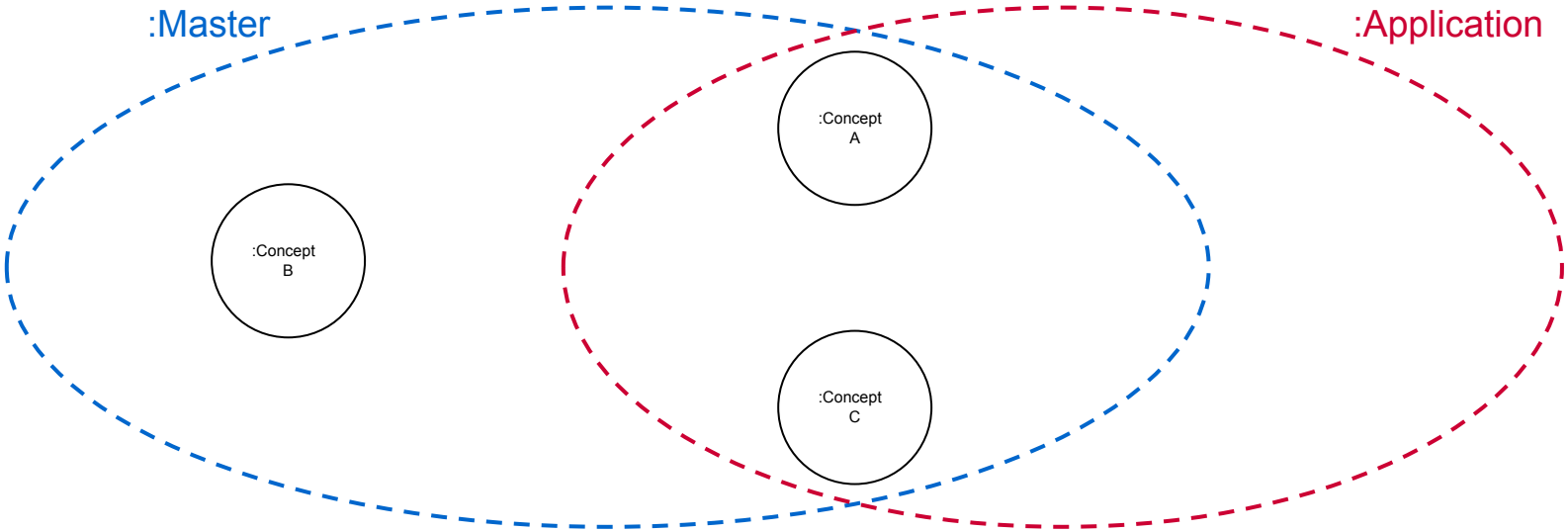
Application Users and Developers

- “I will only use 20 terms and I just want to see those” → **Individual subsets**
- “In our world C is a direct child of A - there is no B in between” → **Individual hierarchies**
- “I want this to be called NSCLC. We all know what we are talking about.” → **Individual labels**
- “Can you just give us the REST endpoint/DB table that fills our dropdown field - thanks!” → **Easy access**
- “Go to our SPARQL endpoint and execute

```
SELECT * WHERE {
   ?disease rdf:type skos:Concept .
   ?disease skos:inScheme roche:Indication .
   ?disease skos:prefLabel ?label .
}
```

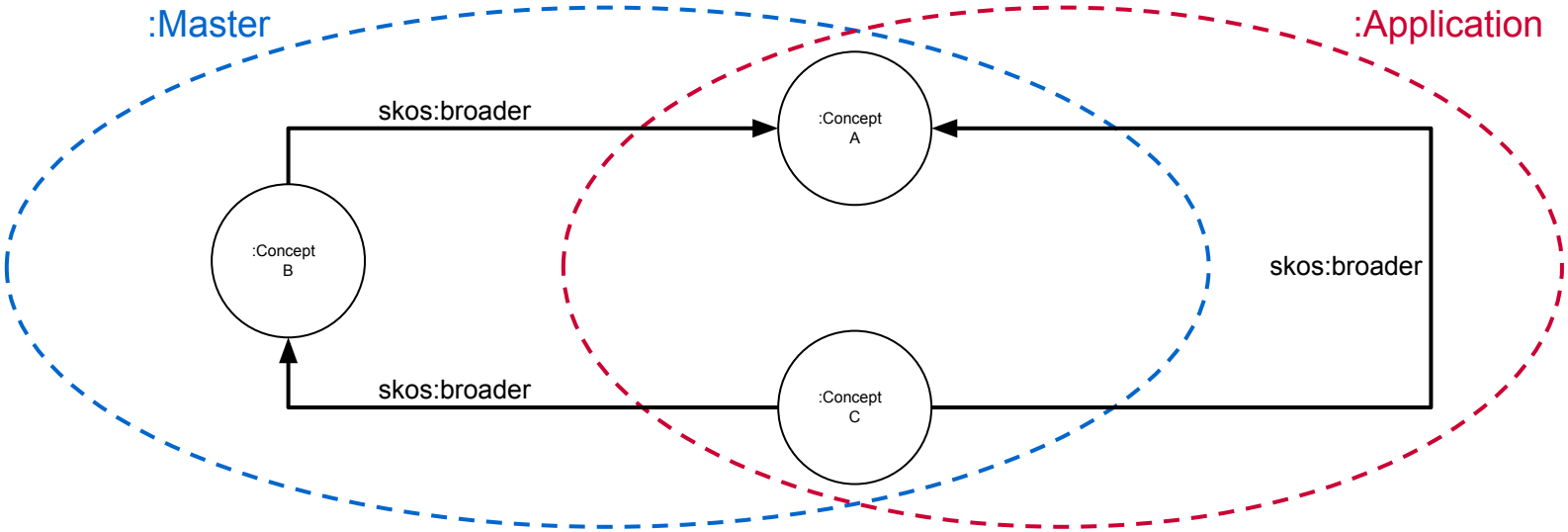
You will get all diseases of the thesaurus.”

Master vs. Application: Subsets



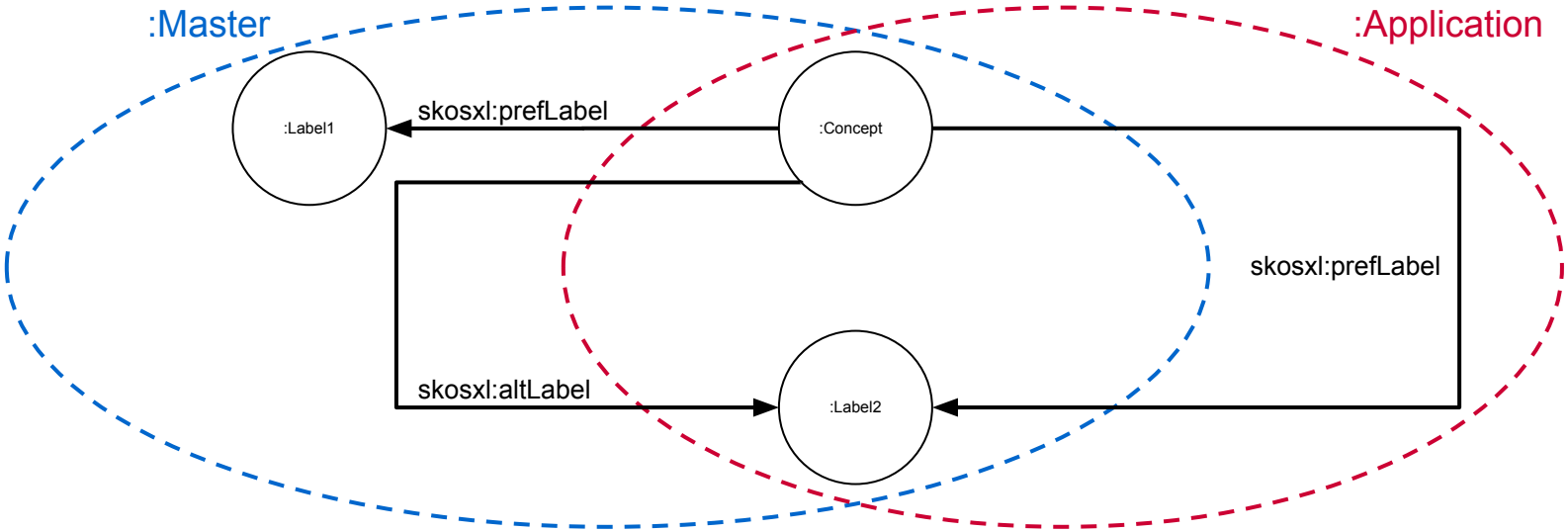
Master Terminology
Application Terminology

Master vs. Application: Hierarchies



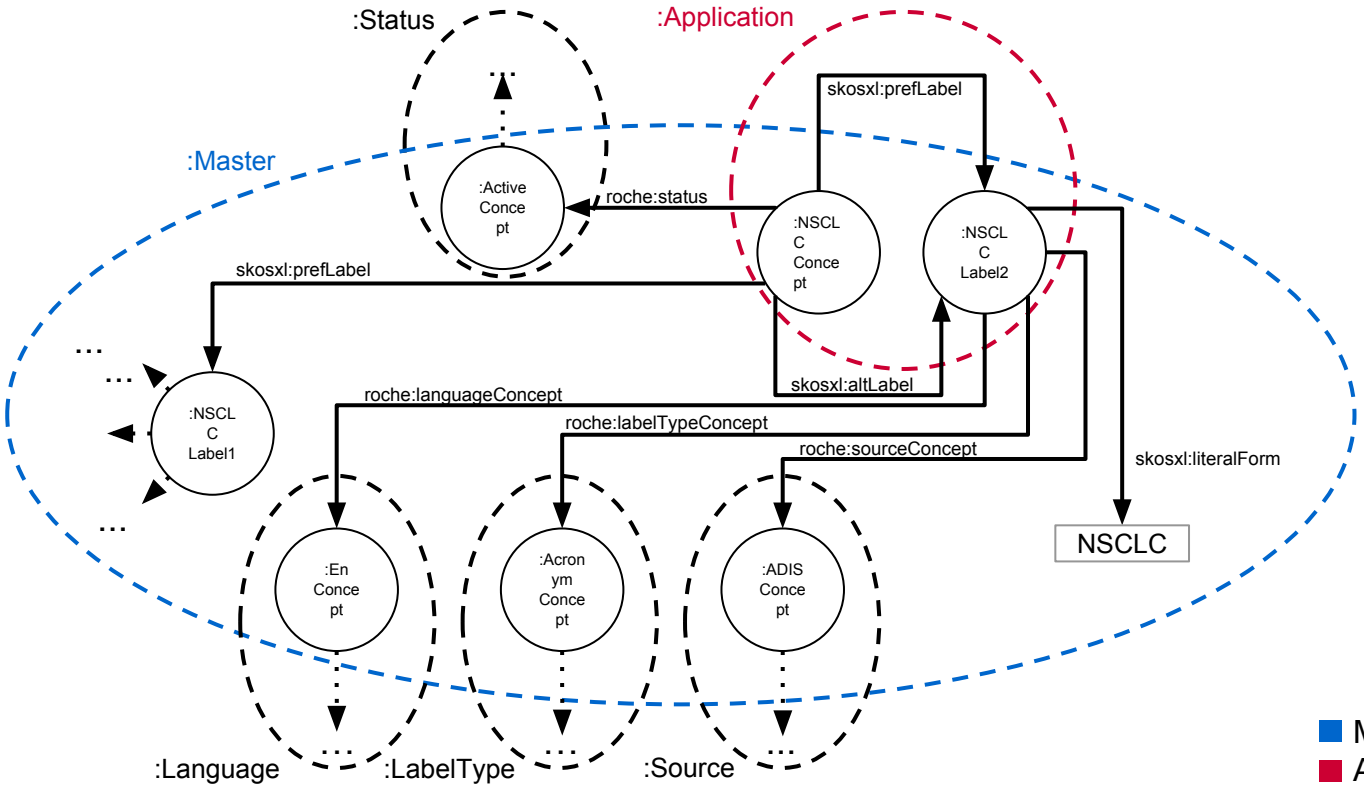
Master Terminology
Application Terminology

Master vs. Application: Preferred Labels



Master Terminology
Application Terminology

Master vs. Application: Complete Picture



- Master Terminology
- Application Terminology
- Meta Terminology

A Headache is a Headache is a Headache...

Phenotype

Sign & Symptom

Side Effect

Indication

Disease

Adverse Event

Exclusion Criterion

Inclusion Criterion

Slide courtesy of Martin Romacker

Semantic Interpretation

Semantic Roles in Application Contexts

Lean Approach to Terminology Management:
Applications define ***contexts***, contextual roles are captured by subsets

Roche Clarity 3PM

Project Partner

Licensor Licensee

Competitor Information Access Platform

Company
[12th Man Technologies ... Abion] - (100)
12th Man Technologies
13Therapeutics
180 Therapeutics
1abtik
1Globe Biomedical
1st Order Pharmaceuticals
2-BBB Medicines
20/20 Gene Systems
21st Century Therapeutics
22nd Century Group
23andMe
2A Pharma AB
2cureX
2M Companies
2X Oncology
3-D Matrix
3-V Biosciences
302 Military Hospital of China
3B Pharmaceuticals
3D Matrix
3DMed
3E Therapeutics

Roche Non-Clinical Repository

Vendor Source

CV_VENDOR_SOURCE
Aids Reagent Program
Asterand
ATCC - American Type Culture Collection
BCRC - Bioresource Collection and Research Center
BCRJ - Banco de Celulas do Rio de Janeiro
BD Biosciences
Caliper
Cell Applications
Cell Systems
Cellartis AB
Clontech
CLS - Cell Lines Service
Coriell Institute
Creative Bioarray

Supplier

NePALS

Semantic Role: same concept with ***contextual*** interpretation (same URI ☺)
Semantic integration versus knowledge proliferation (highly scalable)

Slide courtesy of Martin Romacker

The pREDi Terminology System (RTS)

pRED Informatics - Terminology Service v1.3.0 built on 2017-09-06 16:24:29 [PRD]

Terminology | Application | Curation | Administration | Information

andreas.thalhammer@roche.com Logout

Search

Roche

Scope: Company

Fuzzy search: No

Terminology Navigator

- HMNC Brain Health
- Hoag Memorial Hospital
- Hocuslocus
- Hoelzel Diagnostika
- **Hoffmann-La Roche**
- Hollenbeck Pharmaceuti
- Holmdel Pharmaceutical
- Hologic
- Holostem Terapie Avanz
- HolsboerMaschmeyer N
- Holtech Medical
- Holy Stone Healthcare
- HomeoTherapy
- Homology Medicines
- Hong Kong Innovation a
- Hong Kong Jockey Club I
- [Hong Kong Life Sciences an
- [Hunan Jiuzhitang-Siqi Bioph

Application Navigator

- Covalab
- Creative Biolabs
- Dako
- DCS Innovative Diagnost
- Dianova
- eBioscience
- Epitomics
- Evitria
- FarmaMondo
- Fitzgerald Industries Inte
- Genentech
- Genetex
- Genscript
- Glycart Biotechnology
- Hoelzel
- **Hoffmann-La Roche**
- Icosagen
- ImmuSmol
- Jackson
- Leica Biosystems
- Life Technologies
- LifeSpan BioSciences
- Link Genomics
- Merck Millipore
- Merck Sharp and Dohme
- Miltenyi Biotec

Concept Entity Properties

General information Edit

Label: Hoffmann-La Roche

Terminology: Company

Status: Active

Identifier: ROX1305279968782

Definition

Comment

Labels | Cross References | Relations | Application References

Label	Language	Source	Label Type	Lexical Type
	All	All	All	All
Hoffmann-La Roche	en	PIP, Roche, TPP	Synonym	prefLabel
ARIUS Research	en	ADIS, PIP	Synonym	altLabel
ARIUS Research Inc	en	TPP	Synonym	altLabel
Boehringer Mannheim (Schweiz) AG	en	TPP	Synonym	altLabel
Boehringer Mannheim Corp	en	TPP	Synonym	altLabel
Boehringer Mannheim GmbH	en	TPP	Synonym	altLabel

Acknowledgements

<<Project Lead>>

Martin Romacker
Gunther Dörnen

pREDi Solution Delivery & Architecture

<<Project Development Team Member>>

Grzegorz Strawa,
Daniel Pytel, Bartosz Zak,
Adam Zawada, Karol Bialas
Michal Kolacki, Adrian Rutkowski,
Konrad Borowka, Piotr Koszak

ADMD

<<Project Steering and Support>>

Jörg Schmiedle
Joachim Rupp

pREDi Solution Delivery & Architecture

<<Spatial and Graph Expert>>

Alan Wu
Matthew Perry

Oracle

<<Oracle DB Expert>>

Rui Monteiro, Pablo Noya and
Service Team

pREDi Solution Delivery & Architecture/AE&O

Doing now what patients need next

Backup

Semantic Interpretation

Semantic Roles: Contextual Meaning

Query results in BioPortal for «malignant melanoma» Matches in 24 Ontologies

Adverse Event

Malignant melanoma - Medical Dictionary for Regulatory Activities (MEDDRA)
http://purl.bioontology.org/ontology/MEDDRA/10025650
details - visualize - 13 more from this ontology

Genetics

Malignant melanoma - Online Mendelian Inheritance in Man (OMIM)
http://purl.bioontology.org/ontology/OMIM/MTHU001388
details - visualize

Causal

Malignant Melanoma - Cigarette Smoke Exposure Ontology (CSEO)
http://scai.fraunhofer.de/CSEO#Malignant_Melanoma
details - visualize

Phenotype

Melanoma - Human Phenotype Ontology (HP)
http://purl.obolibrary.org/obo/HP_0002861
The presence of a melanoma, a malignant cancer originating from pigment producing melanocytes. Melanoma can originate from the skin or the pigmented layers of the eye (the uvea).
details - visualize

Experimental

melanoma - Experimental Factor Ontology (EFO)
http://www.ebi.ac.uk/efo/EFO_0000756
A malignant neoplasm derived from cells that are capable of forming melanin, which may occur in the skin of any part of the body, in the eye, or, rarely, in the mucous membranes of the genitalia, anus, oral ...
details - visualize

Semantic Role: Same concept with contextual interpretation (different URI ☹)
Knowledge proliferation vs semantic integration

Slide courtesy of Martin Romacker