

Introduction to Linked Open Data for Computational Mass Spectrometry

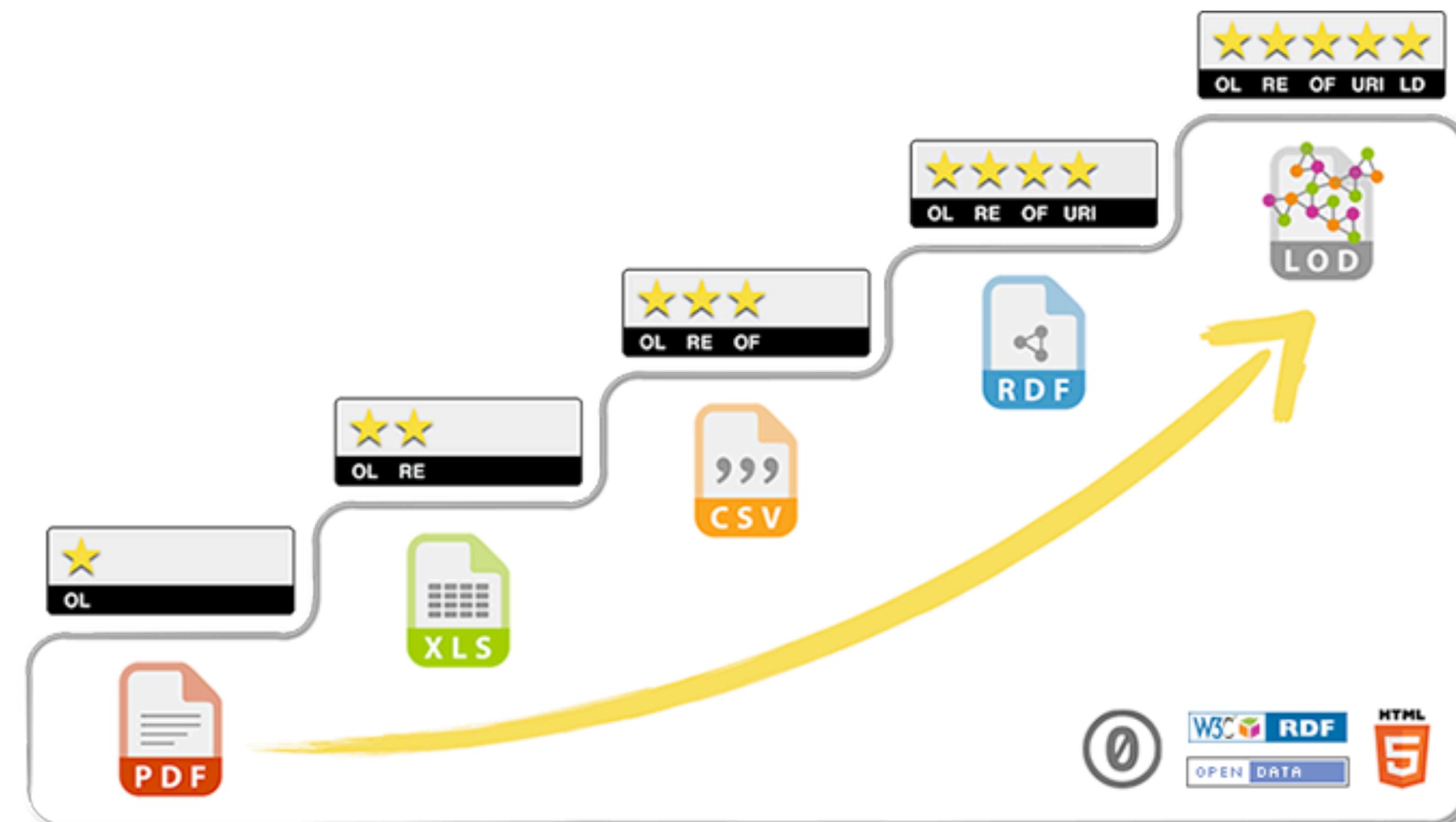
Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences
Flemingovo náměstí 542/2

Adriano Rutz, 2024-04-17



D BIOL

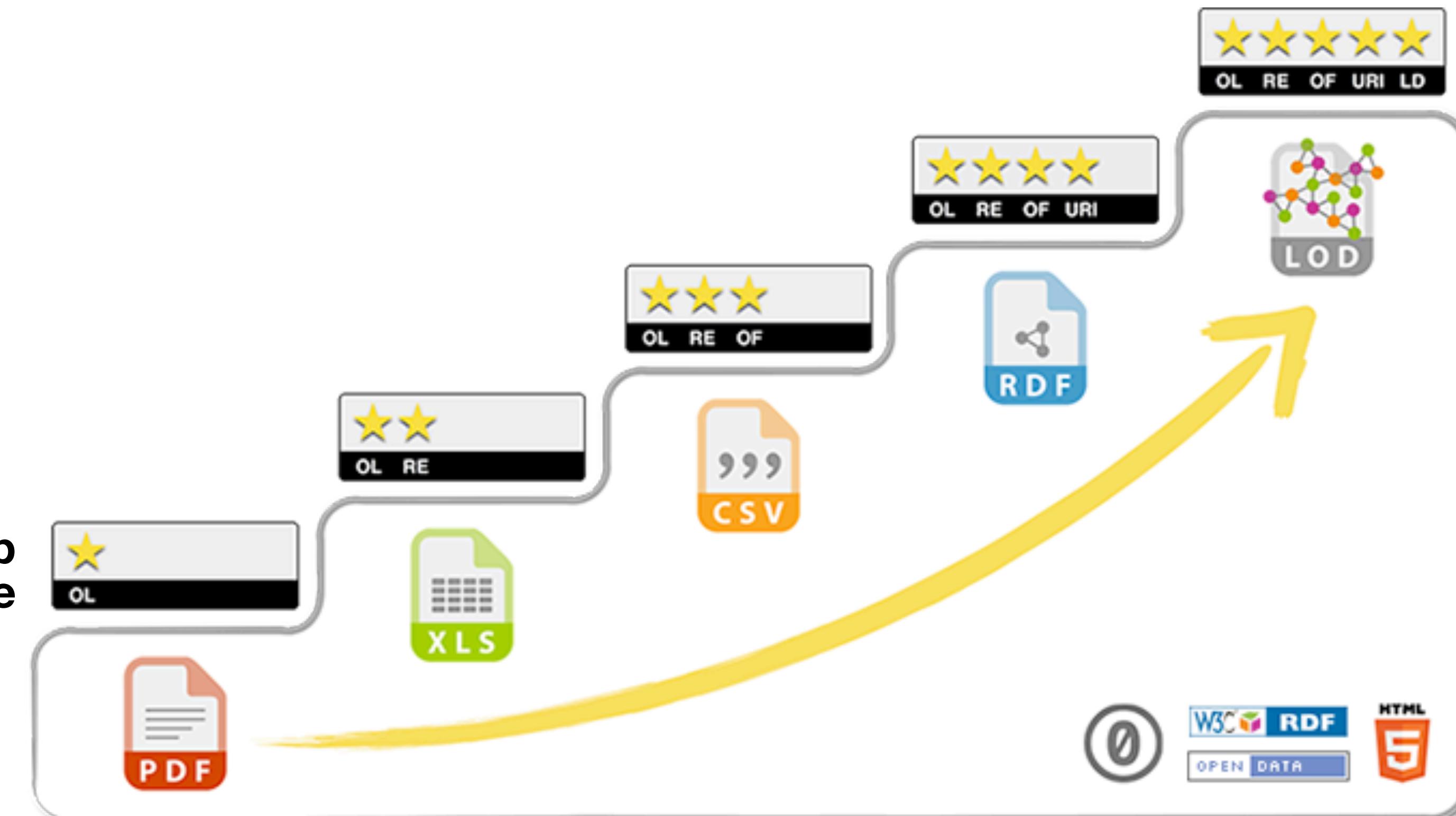
What is Linked Open Data (LOD)?



https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

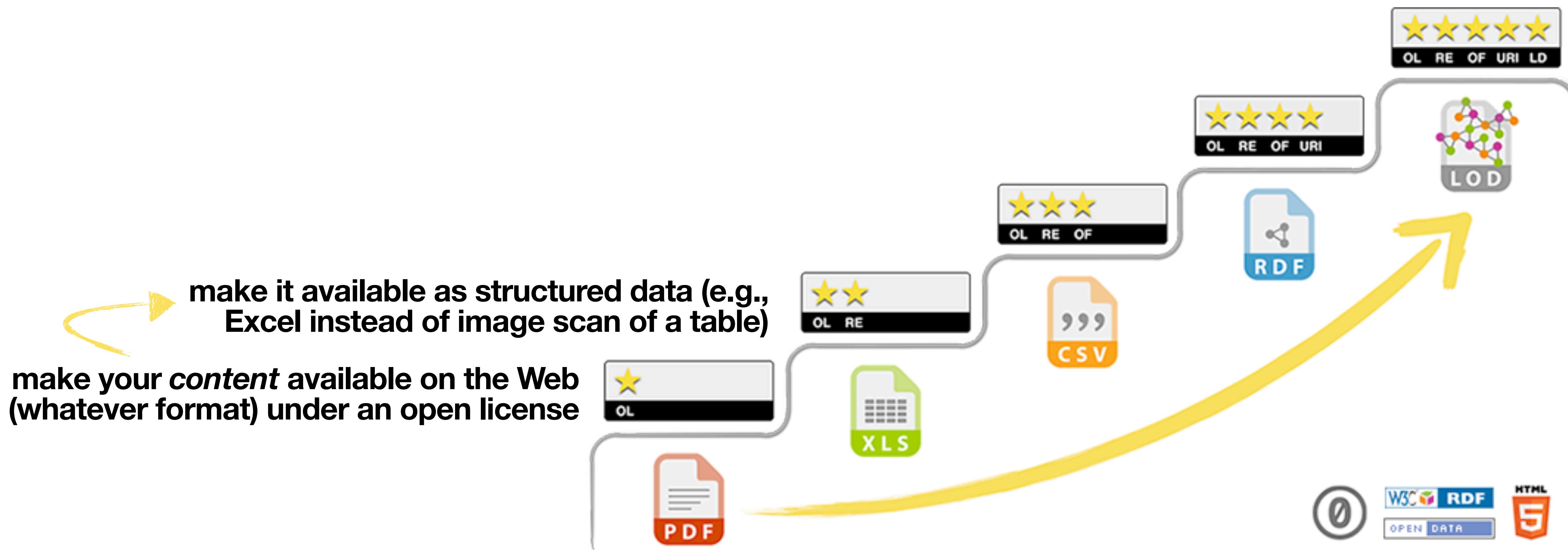
What is Linked Open Data (LOD)?

make your *content* available on the Web
(whatever format) under an open license



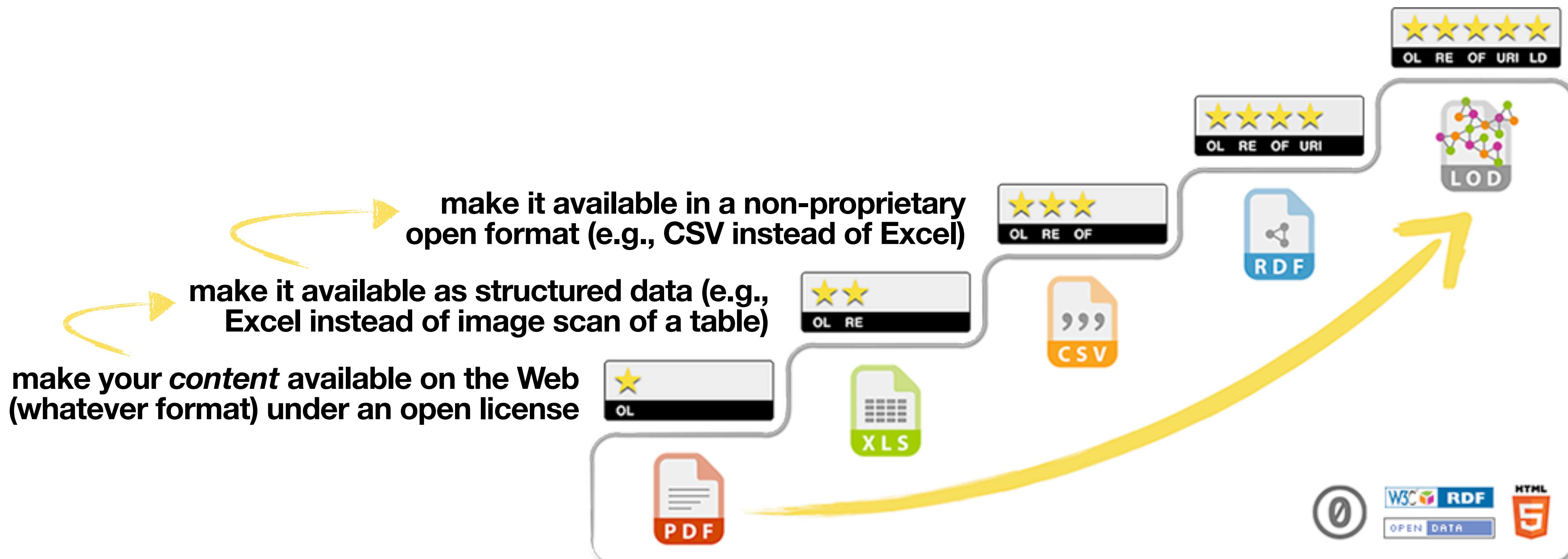
https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

What is Linked Open Data (LOD)?



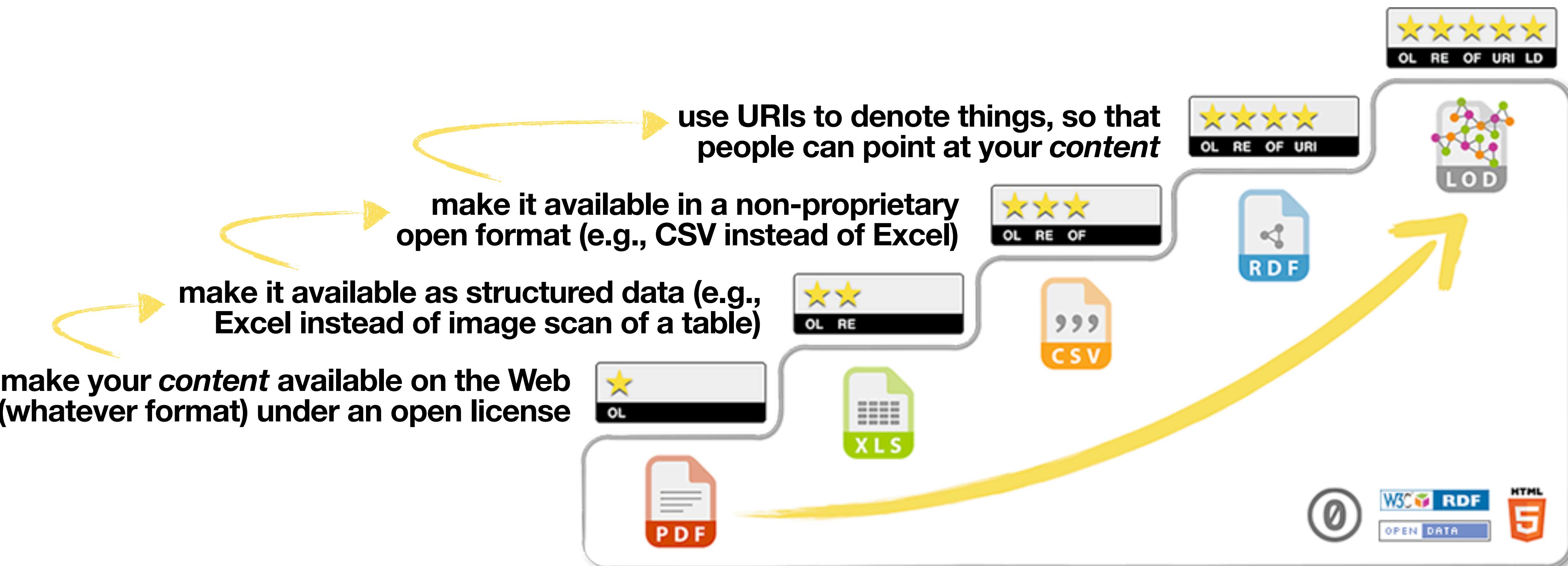
https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

What is Linked Open Data (LOD)?



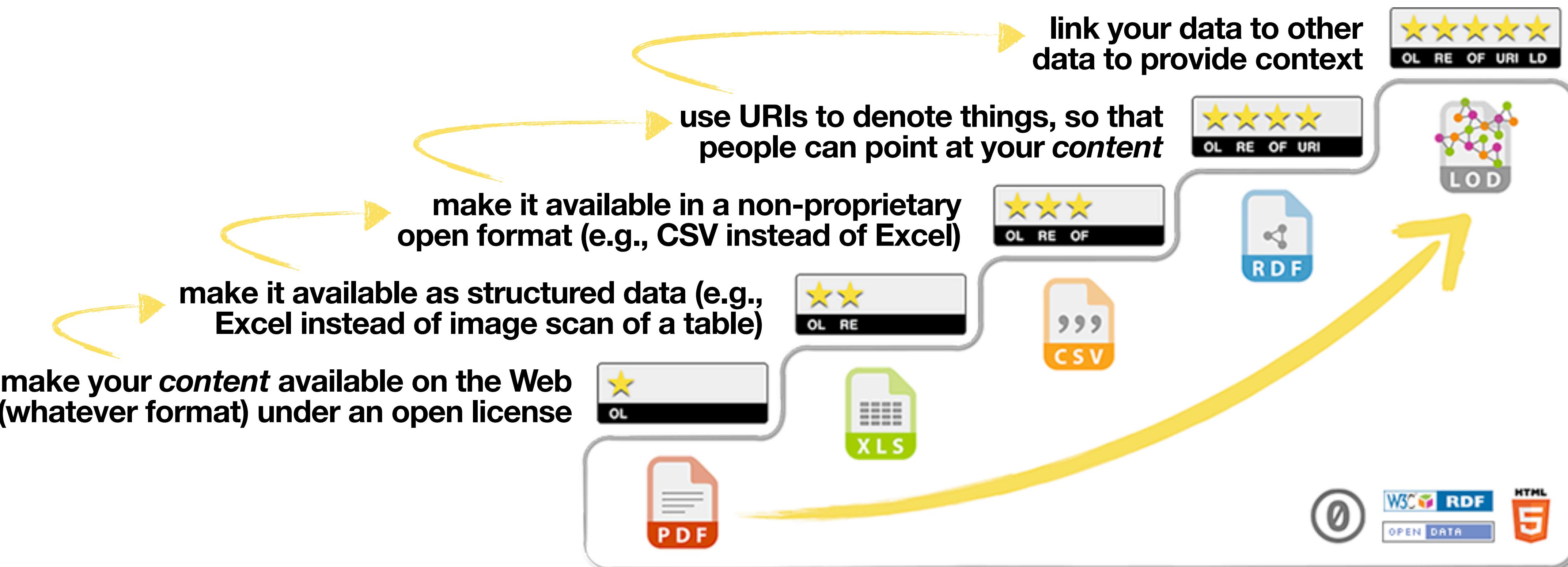
https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

What is Linked Open Data (LOD)?



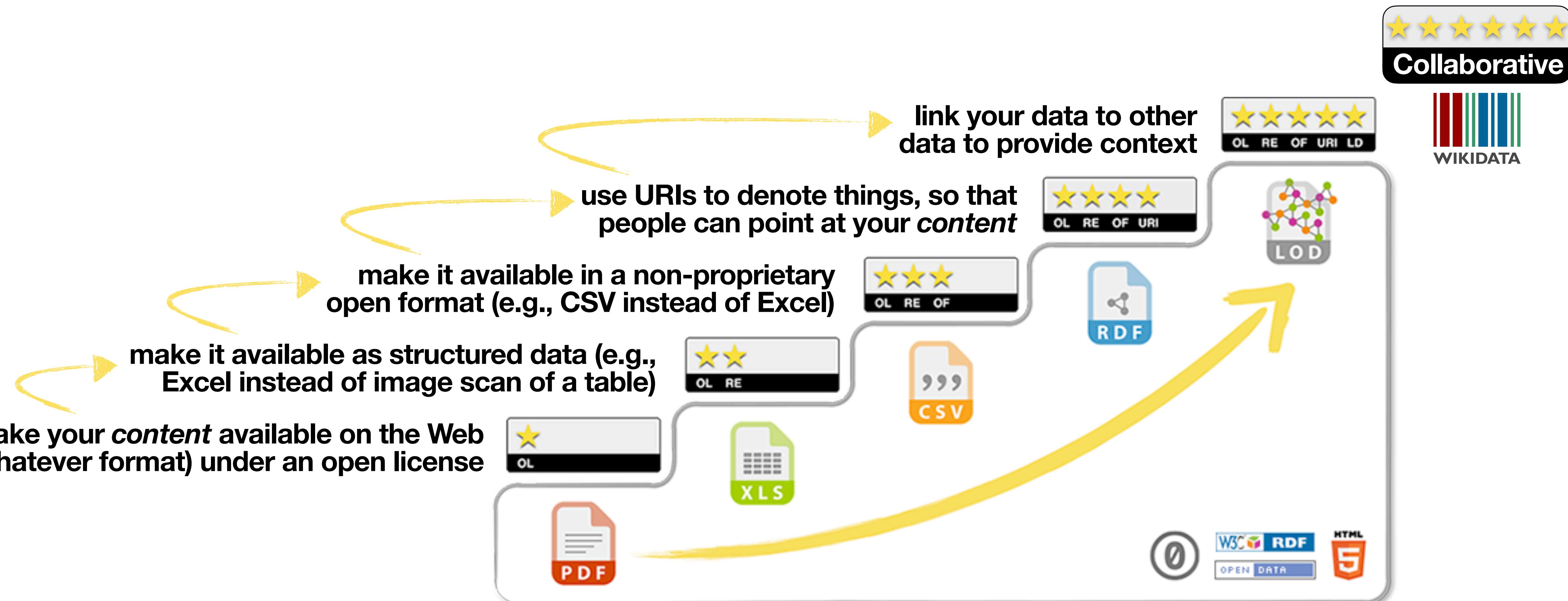
https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

What is Linked Open Data (LOD)?



https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

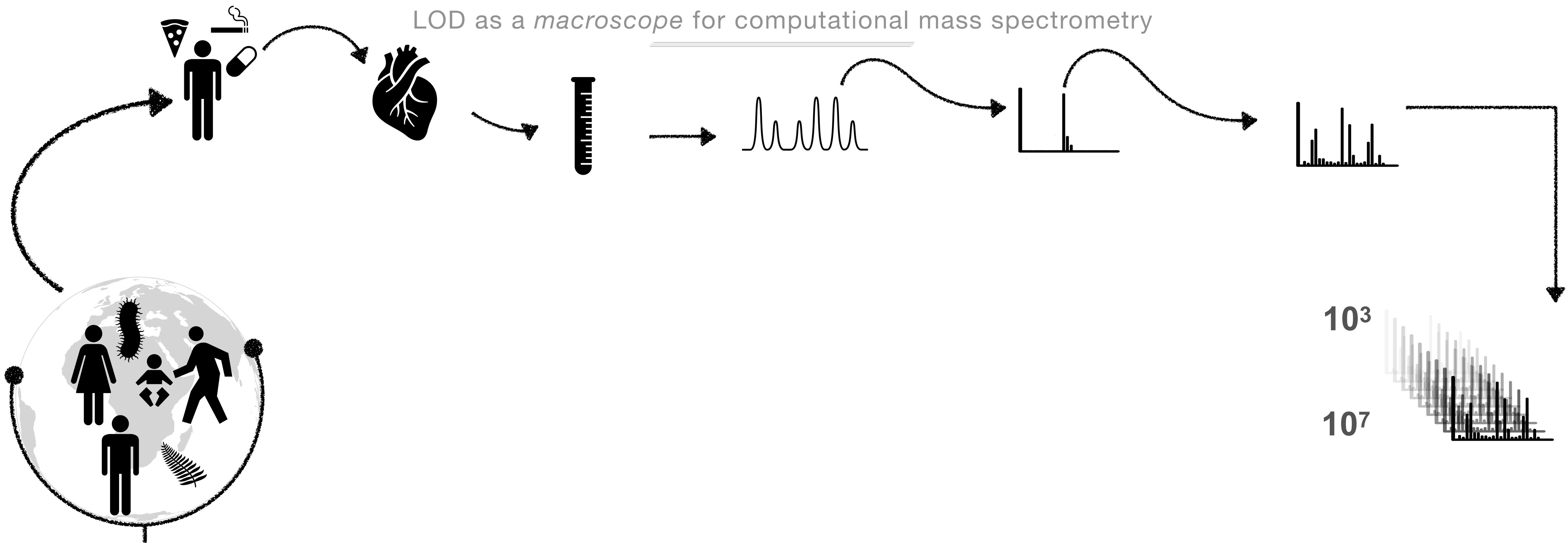
What is Linked Open Data (LOD)?



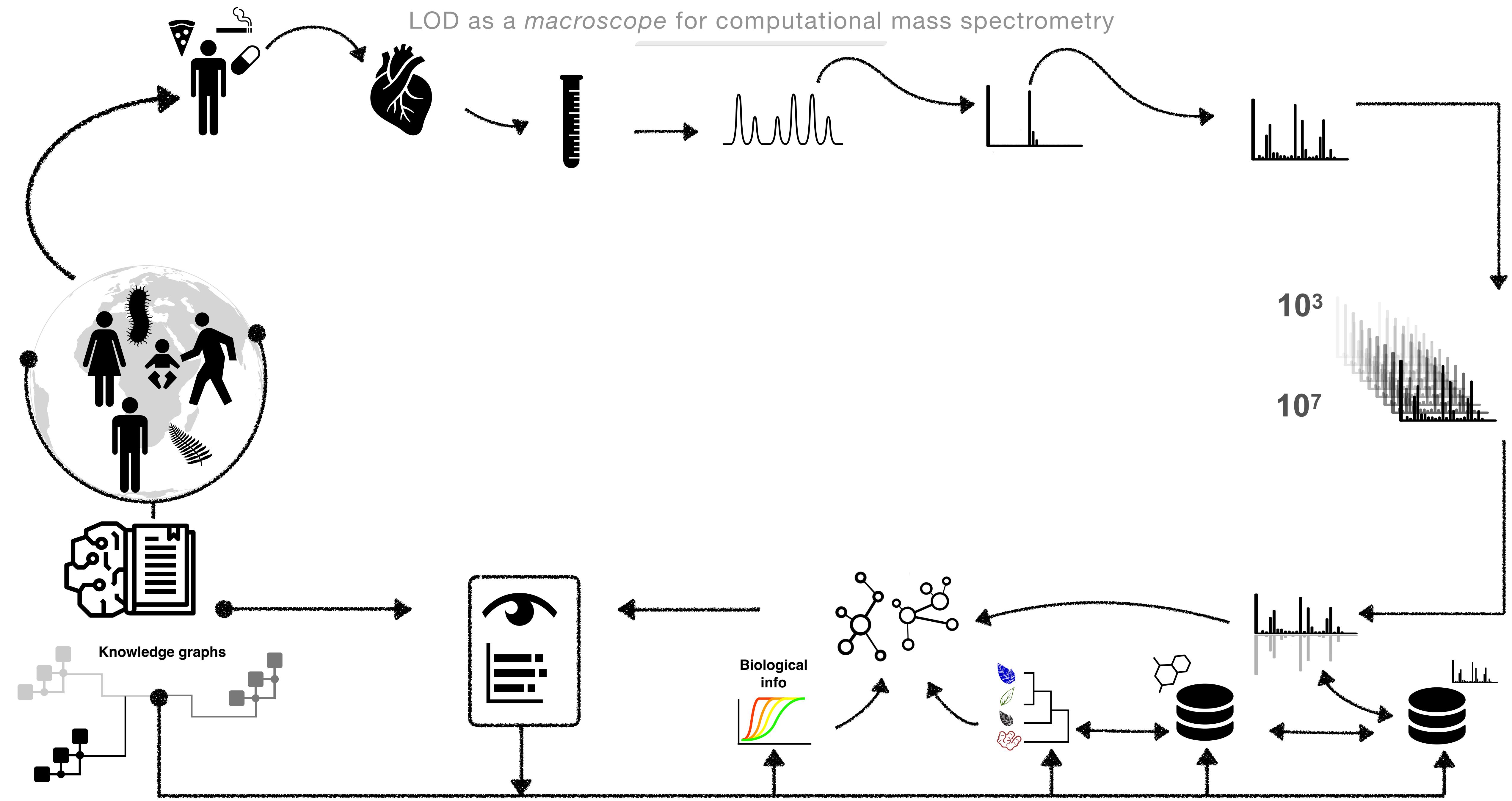
https://commons.wikimedia.org/wiki/File:5-star_deployment_scheme_for_Open_Data.png

LOD as a *macroscope* for computational mass spectrometry

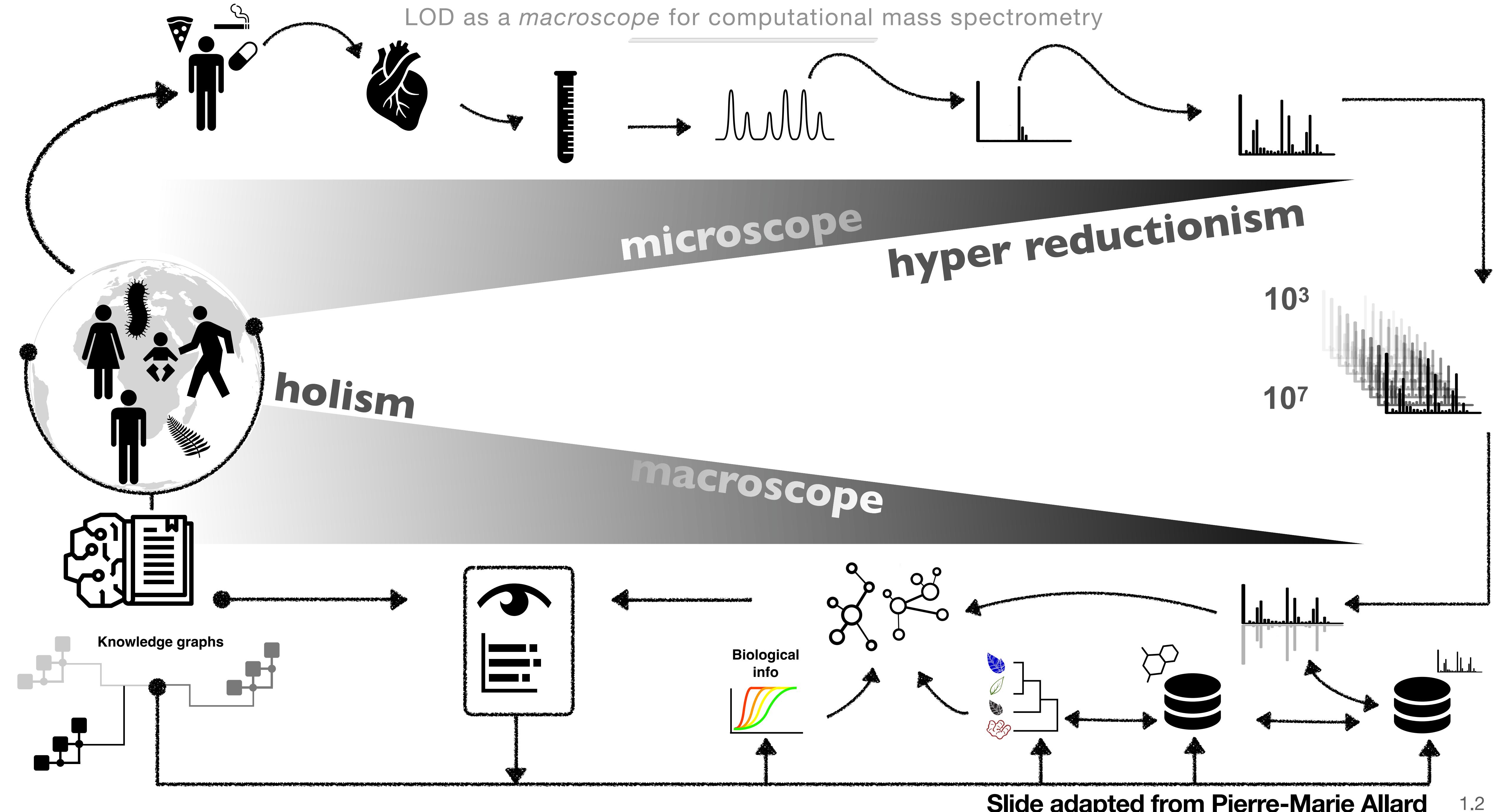




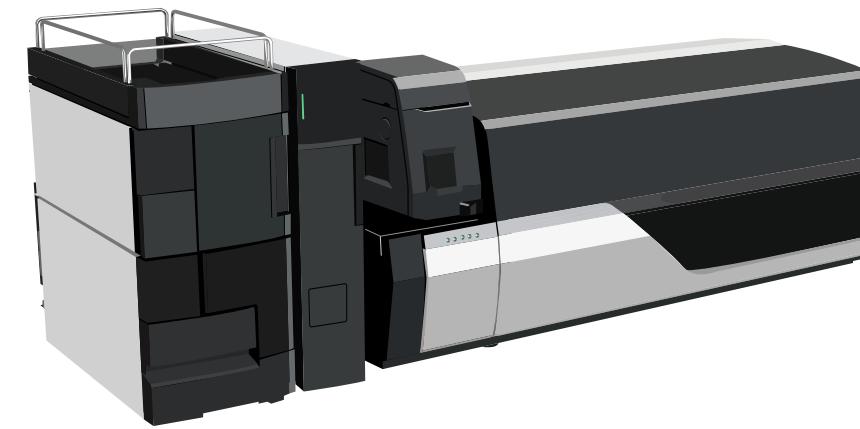
LOD as a macroscope for computational mass spectrometry



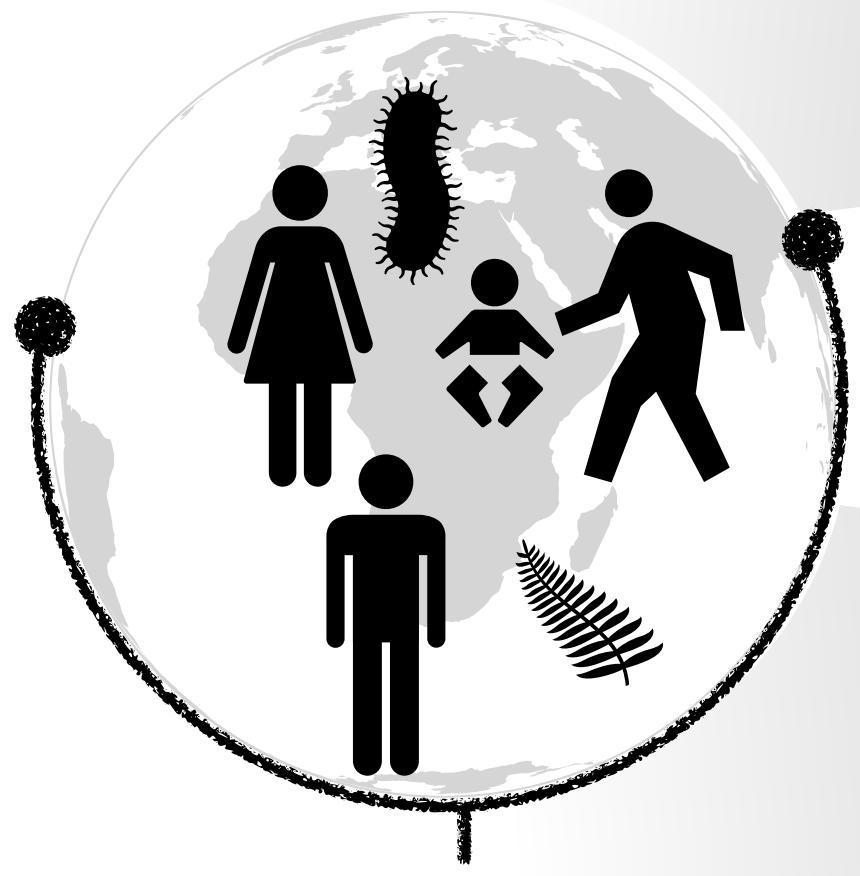
LOD as a macroscope for computational mass spectrometry



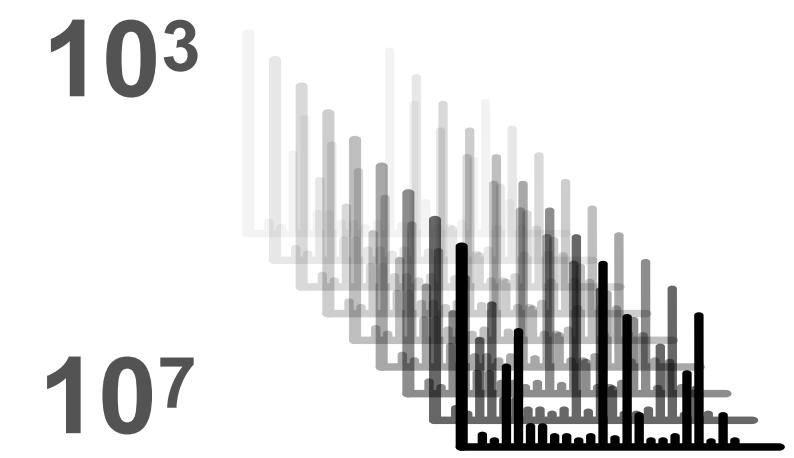
LOD as a *macroscope* for computational mass spectrometry



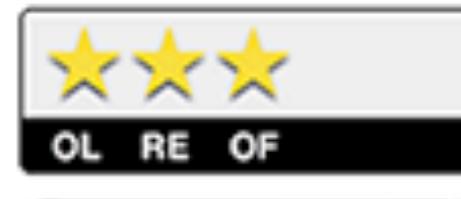
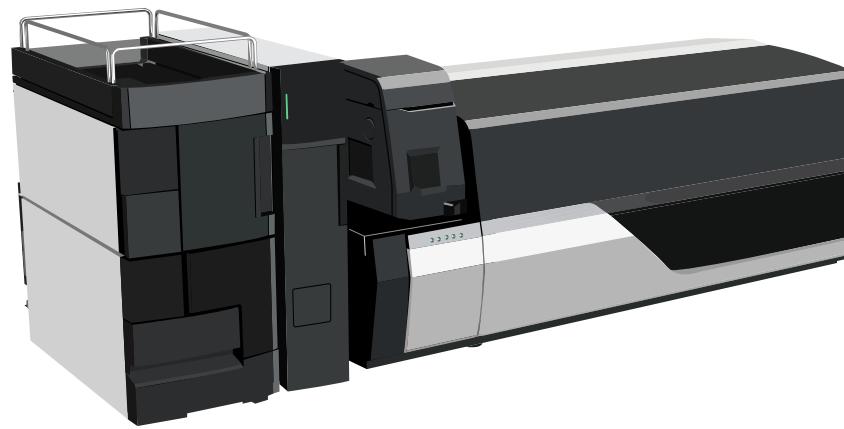
microscope



macroscope

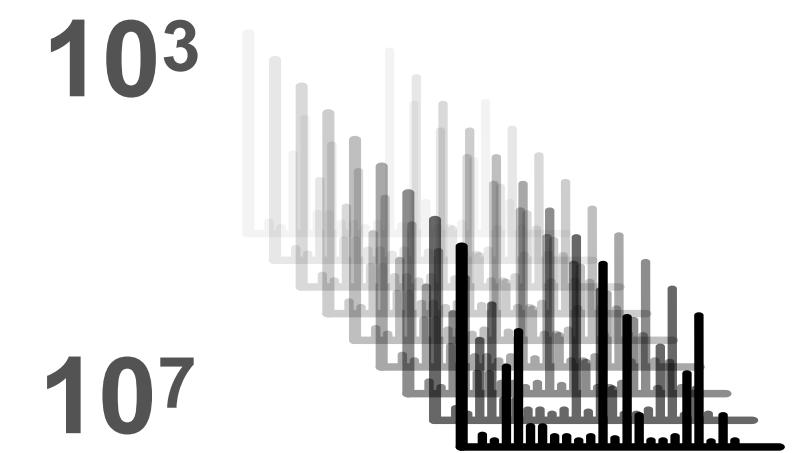
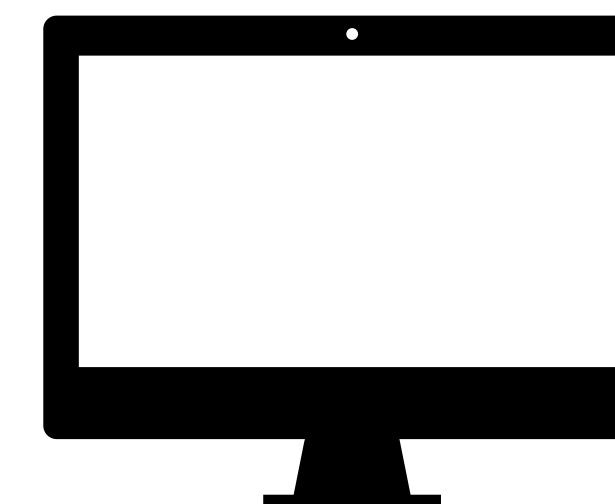
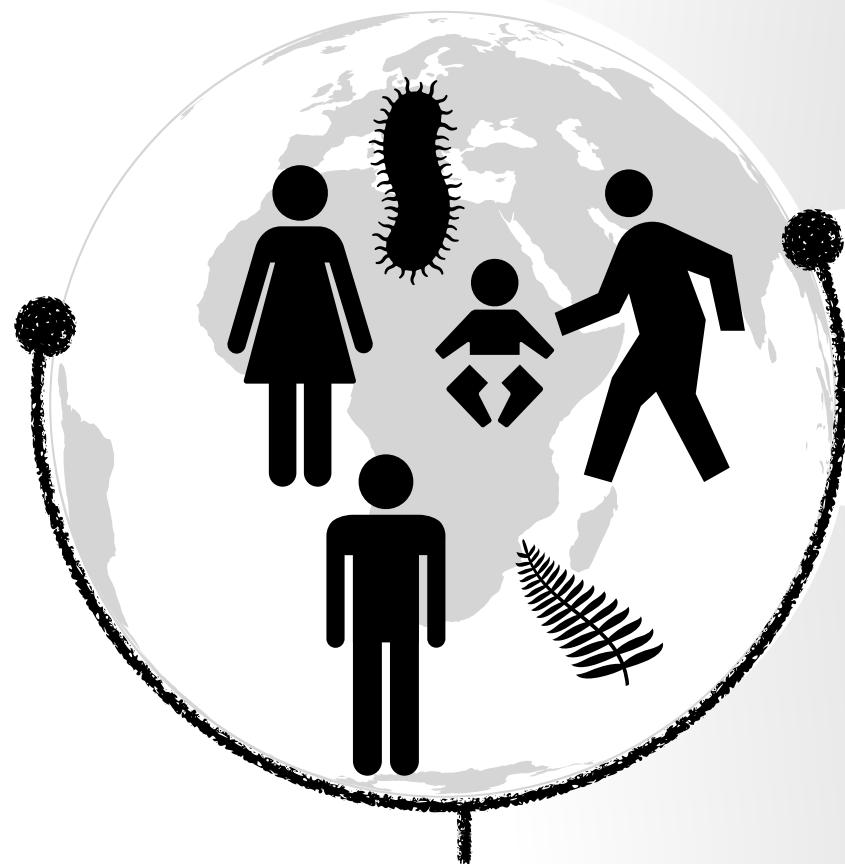


LOD as a *macroscope* for computational mass spectrometry

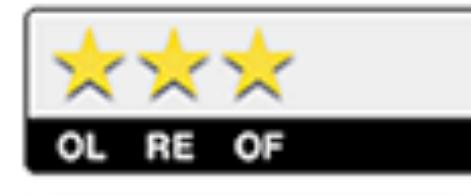
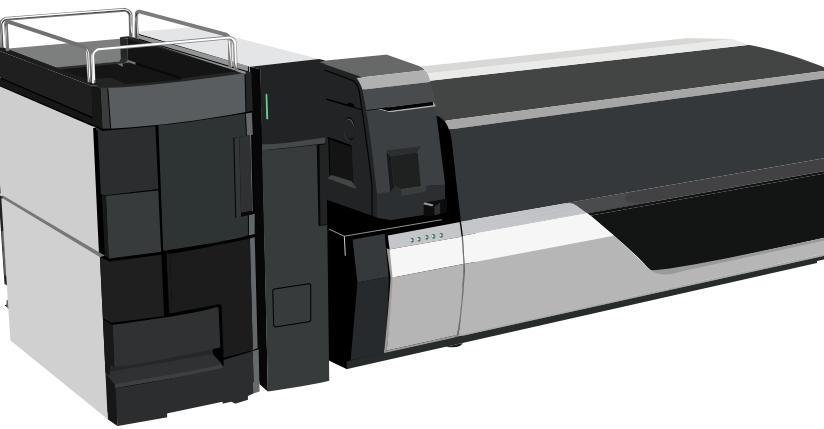


microscope

macroscope



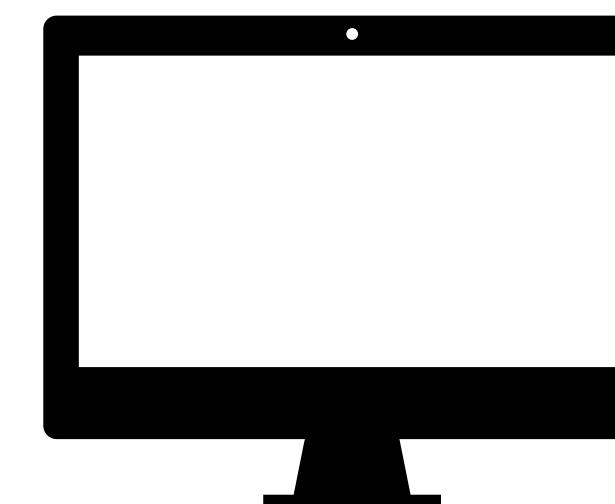
LOD as a *macroscope* for computational mass spectrometry



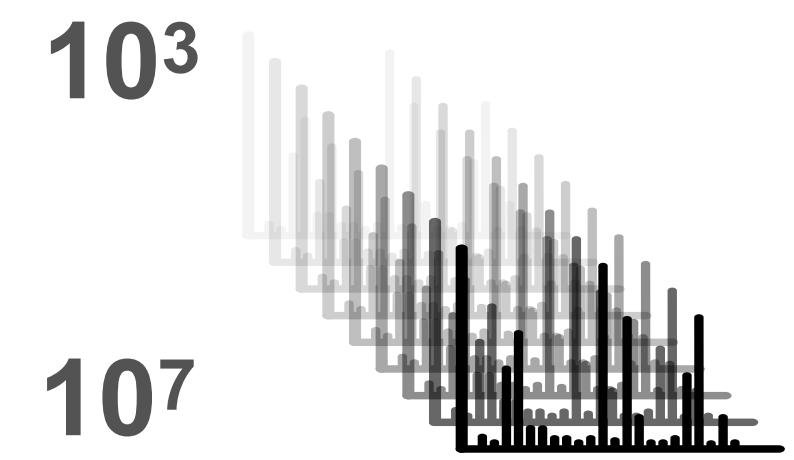
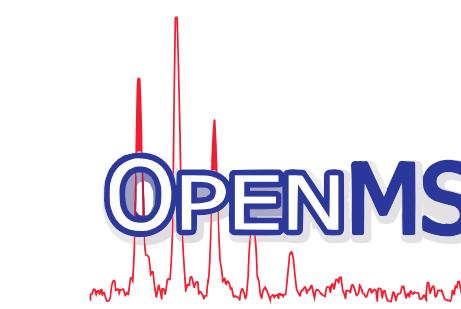
microscope



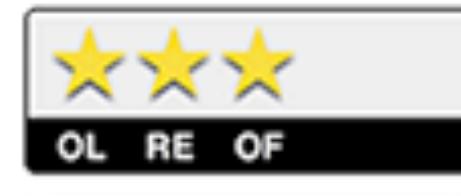
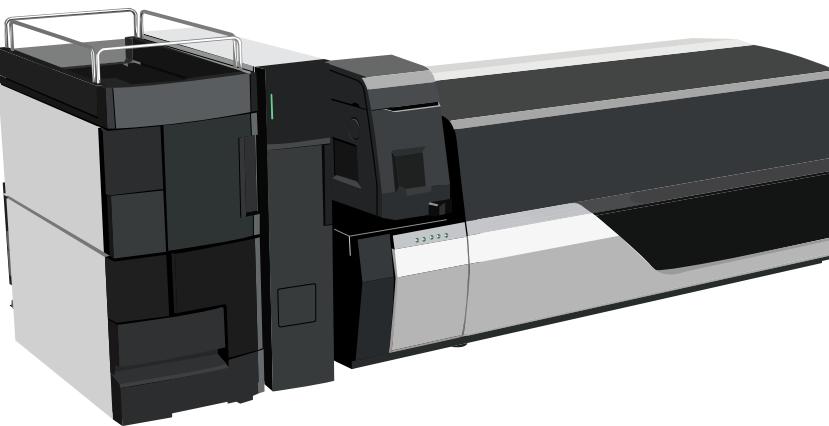
macroscope



mzmine



LOD as a macroscope for computational mass spectrometry



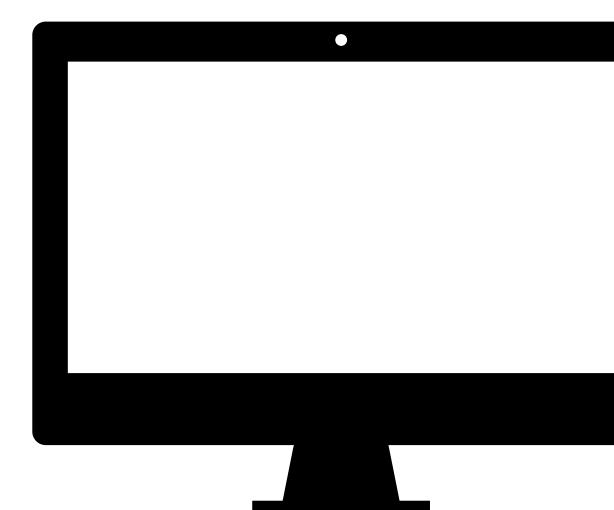
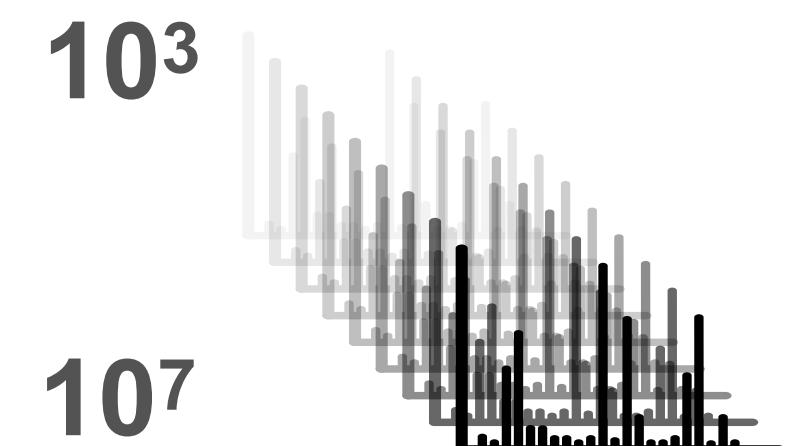
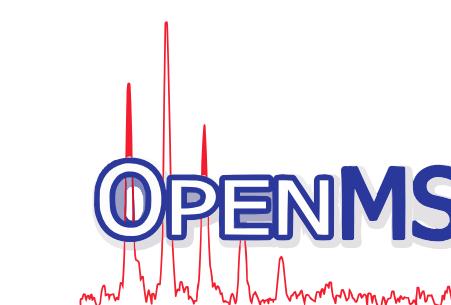
microscope



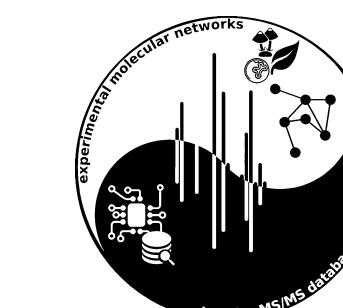
macroscope



mzmine



 MassBank
High Quality Mass Spectral Database



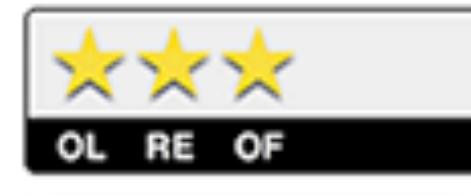
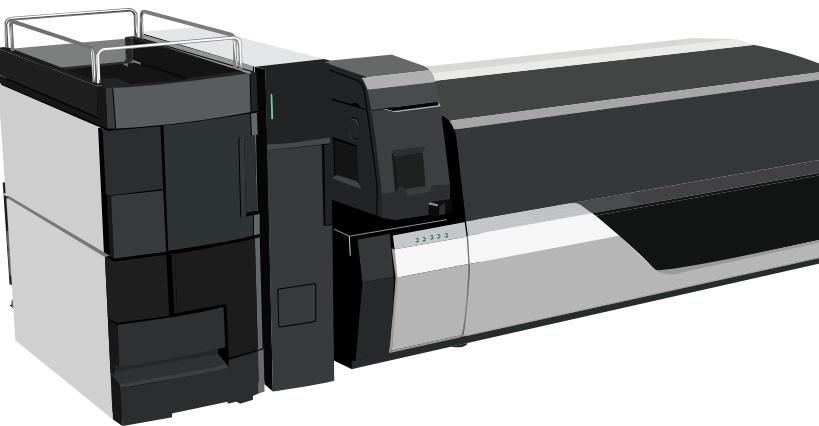
MoNA - MassBank of North America

 MATCHMS

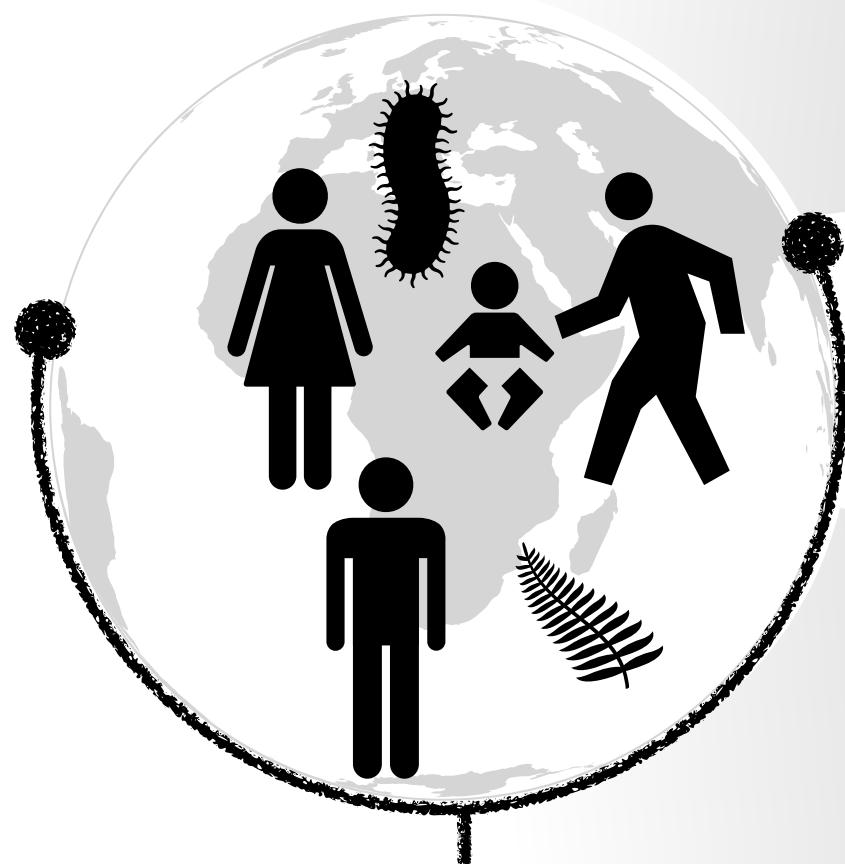
 MSRIUS

 NST

LOD as a macroscope for computational mass spectrometry



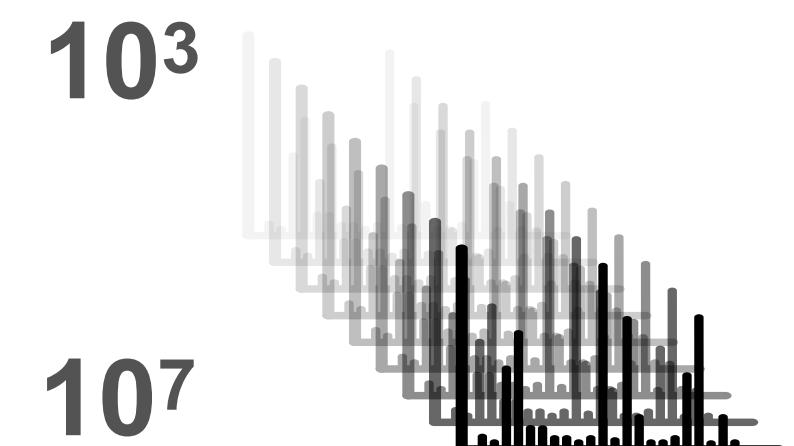
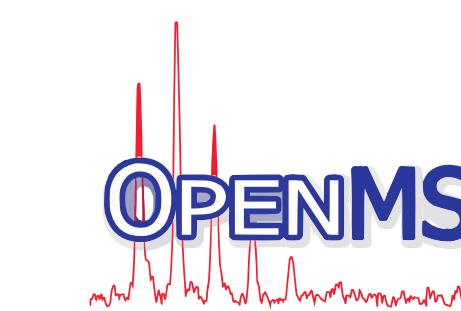
microscope



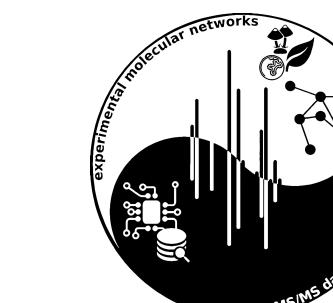
macroscope



mzmine



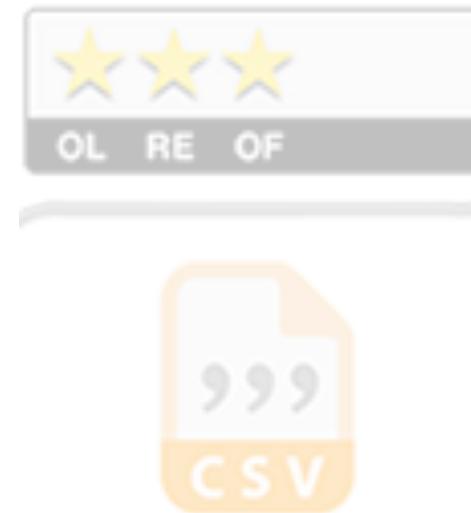
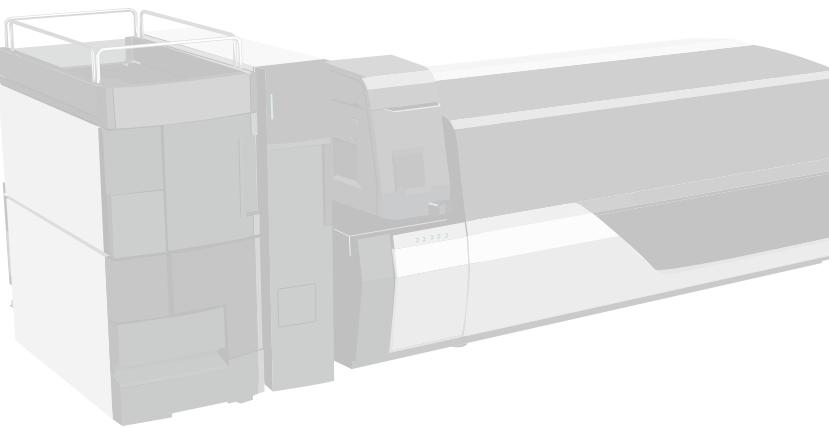
MassBank 
High Quality Mass Spectral Database



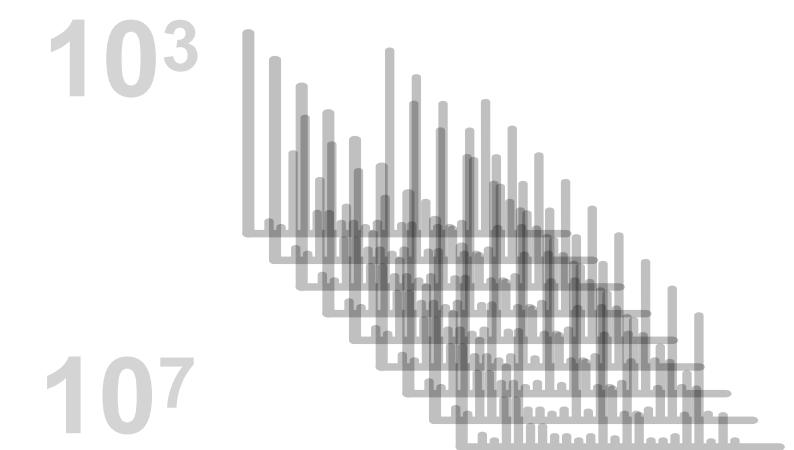
MoNA - MassBank of North America

 MSRIUS
 NST

LOD as a macroscope for computational mass spectrometry



Put your data back into a
broader (biologically
relevant) context!



Wikidata

WIKIPATHWAYS
Pathways for the People

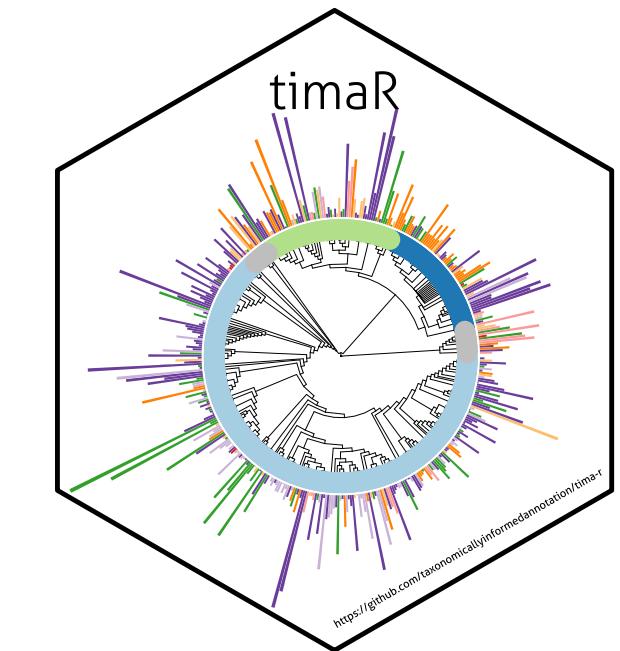
PubChem

GNPS MASST

mass.ucsd.edu

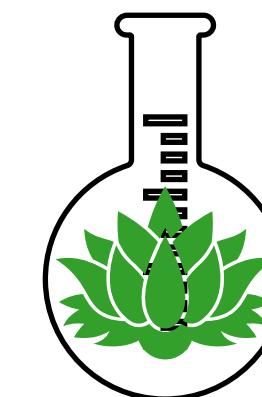


Taxonomically informed metabolite annotation

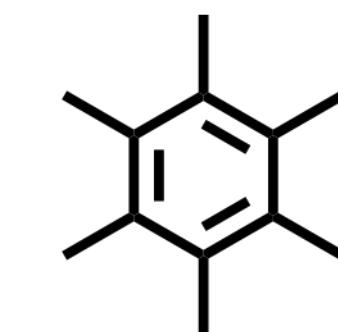


“Plant taxonomy would be the most useful guide to man in his search for new industrial and medicinal plants” (de Candolle, 1804)

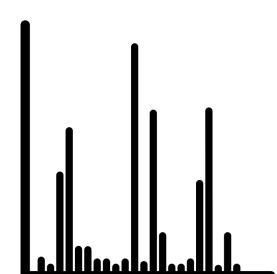
Metabolomics



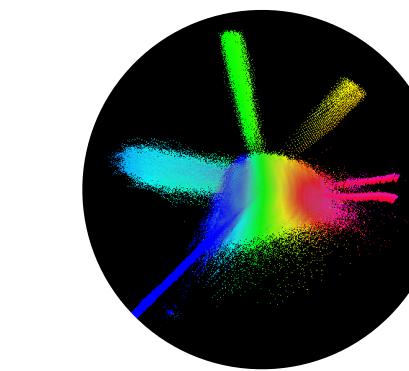
Natural Extract



Structure



MS^2
spectrum

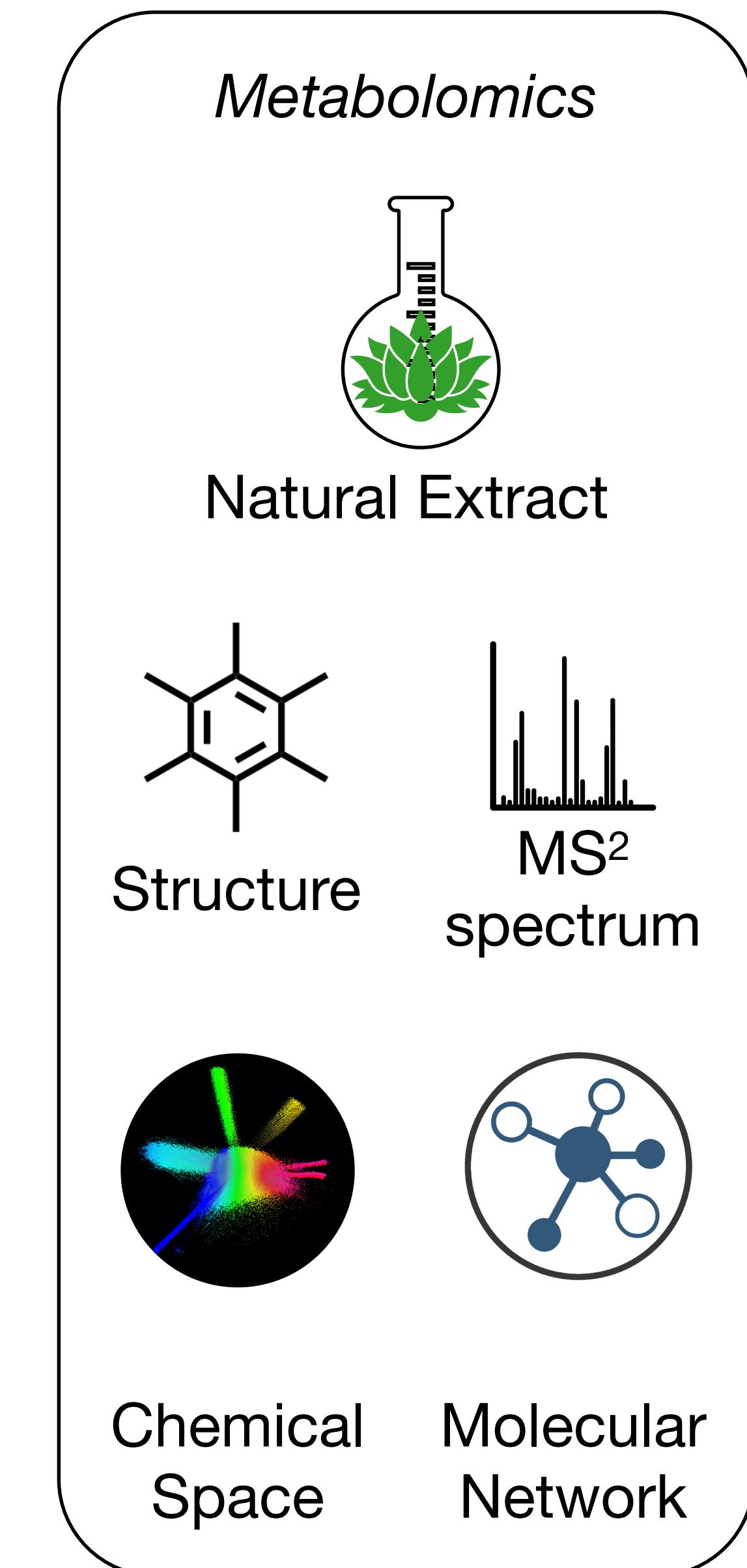
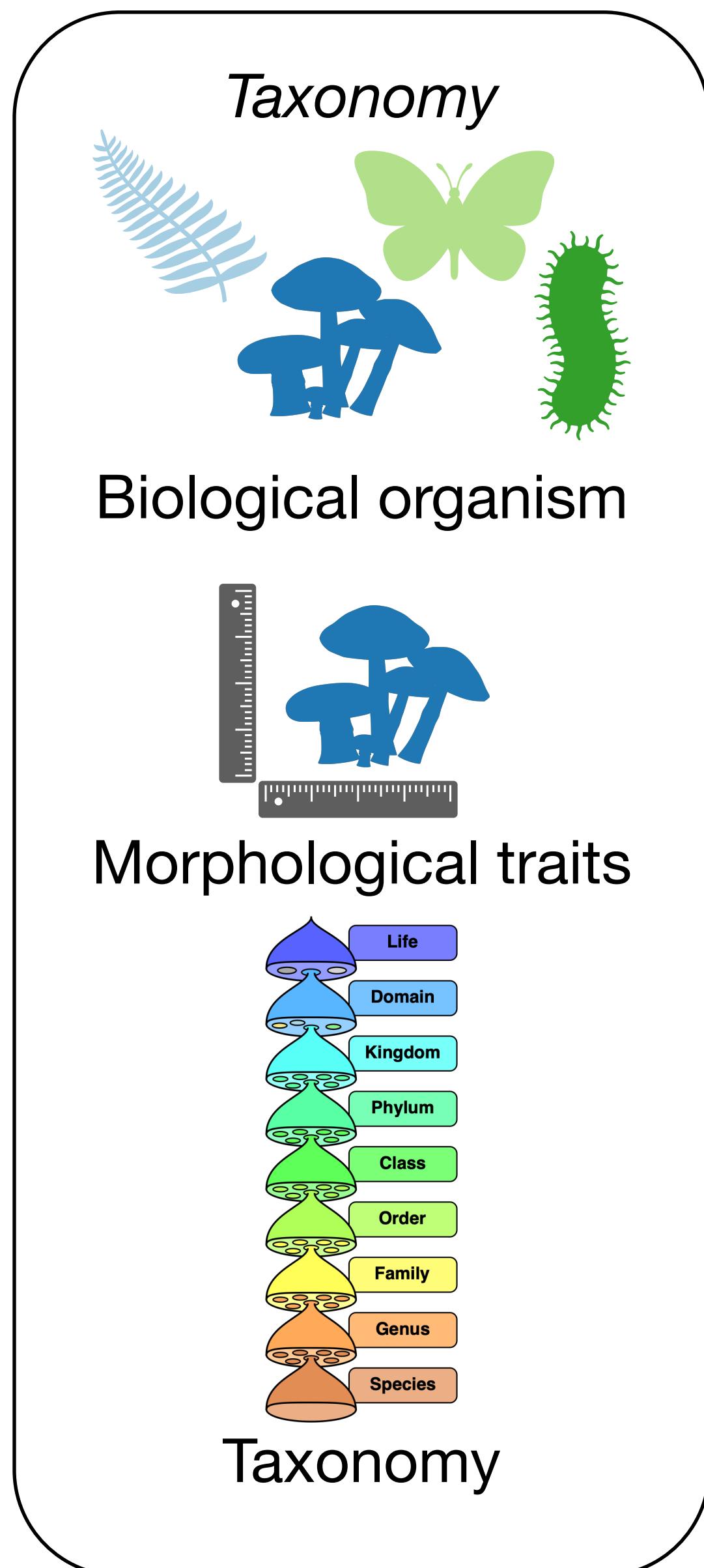


Chemical
Space

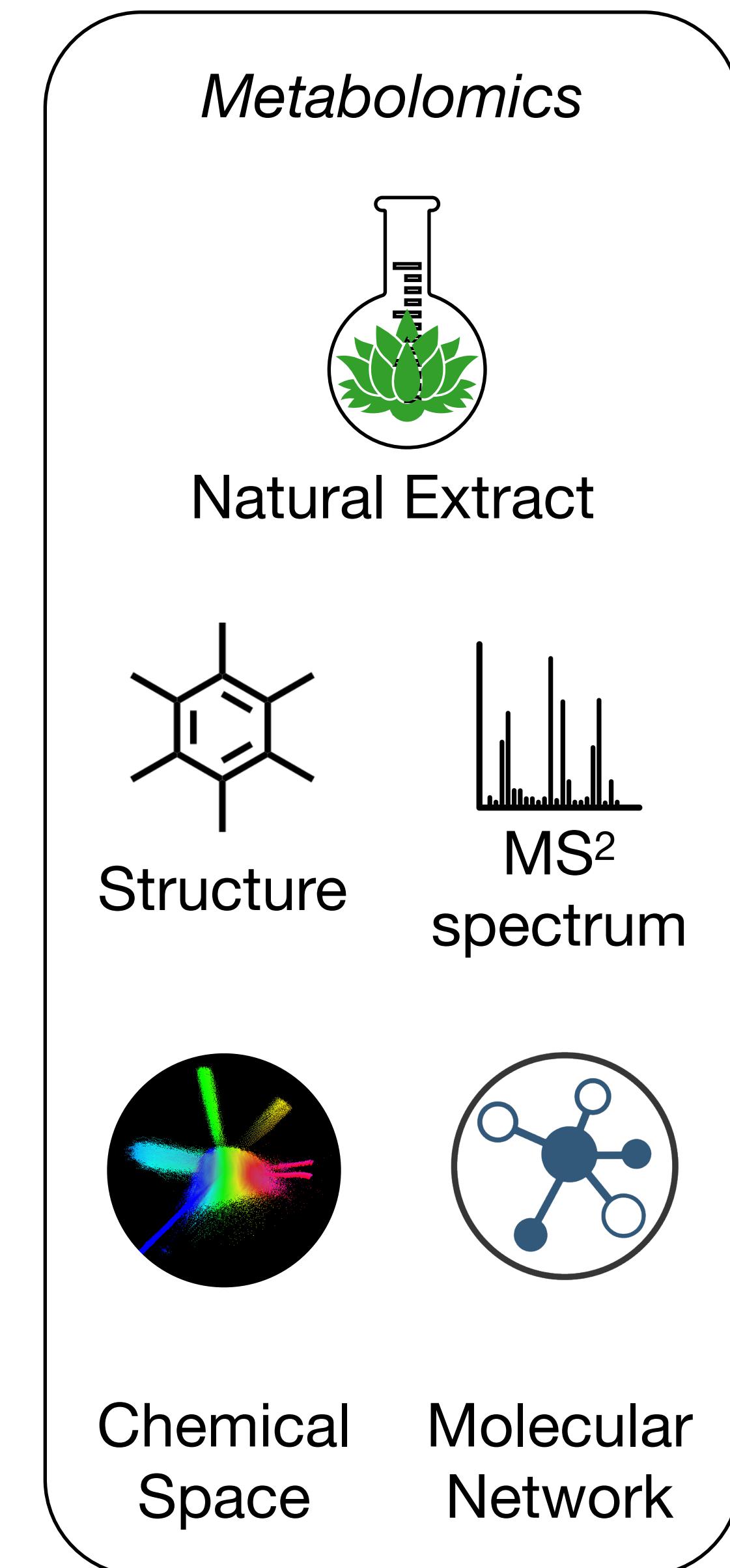
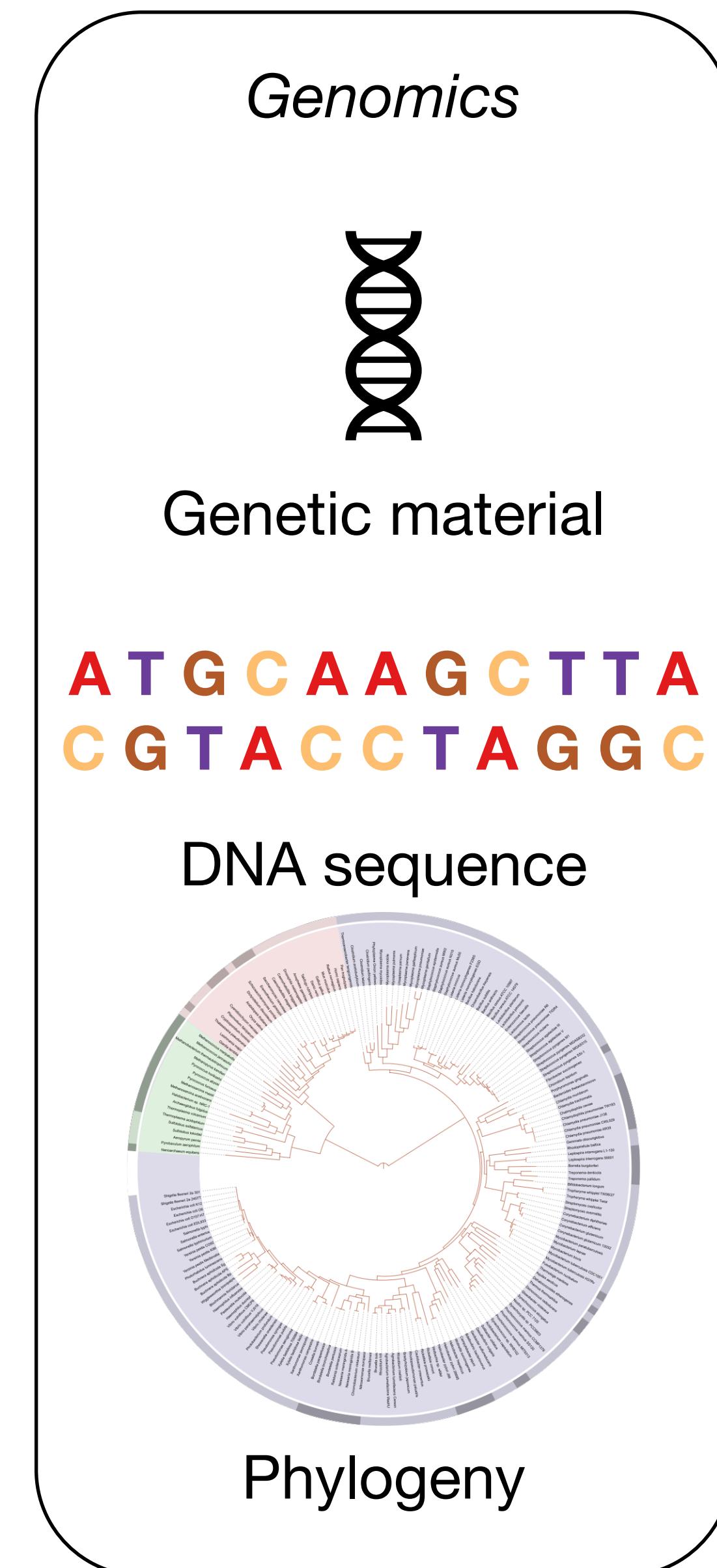
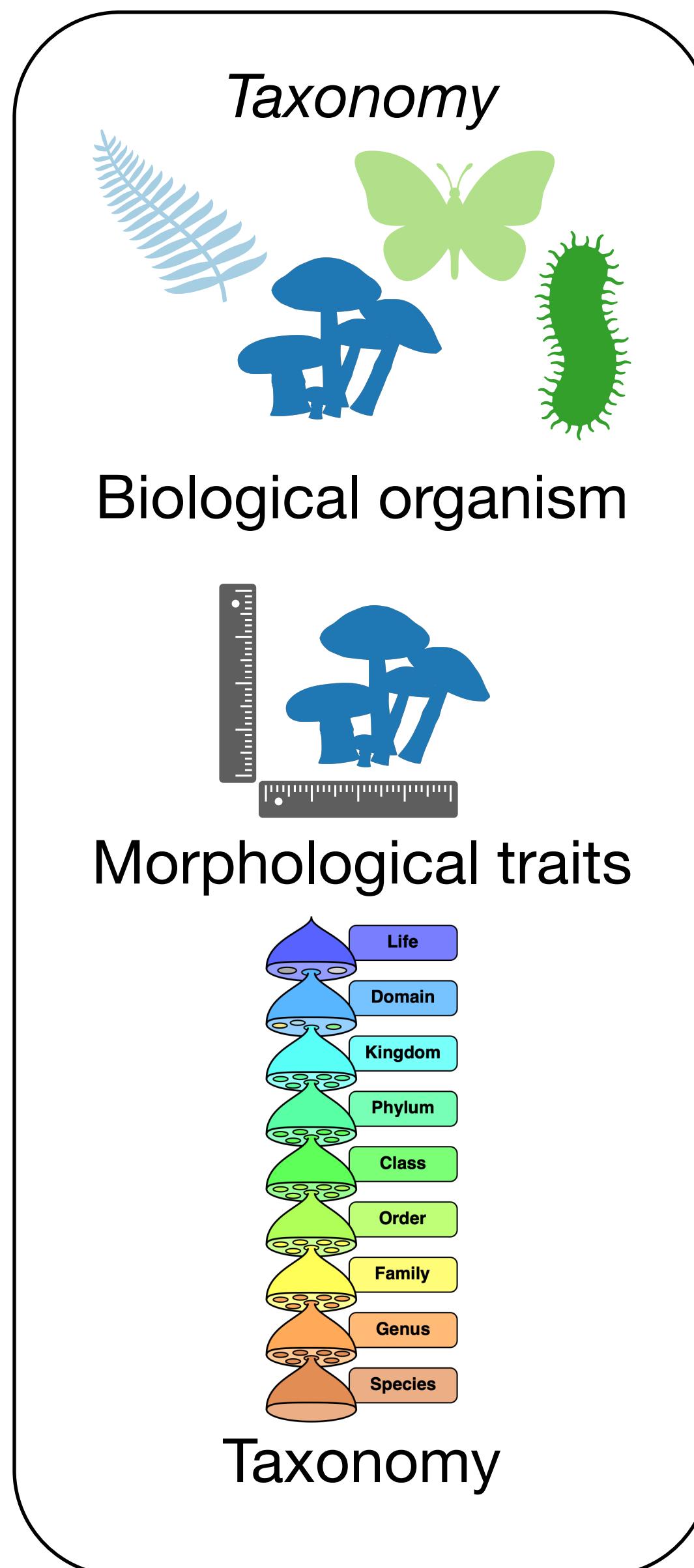


Molecular
Network

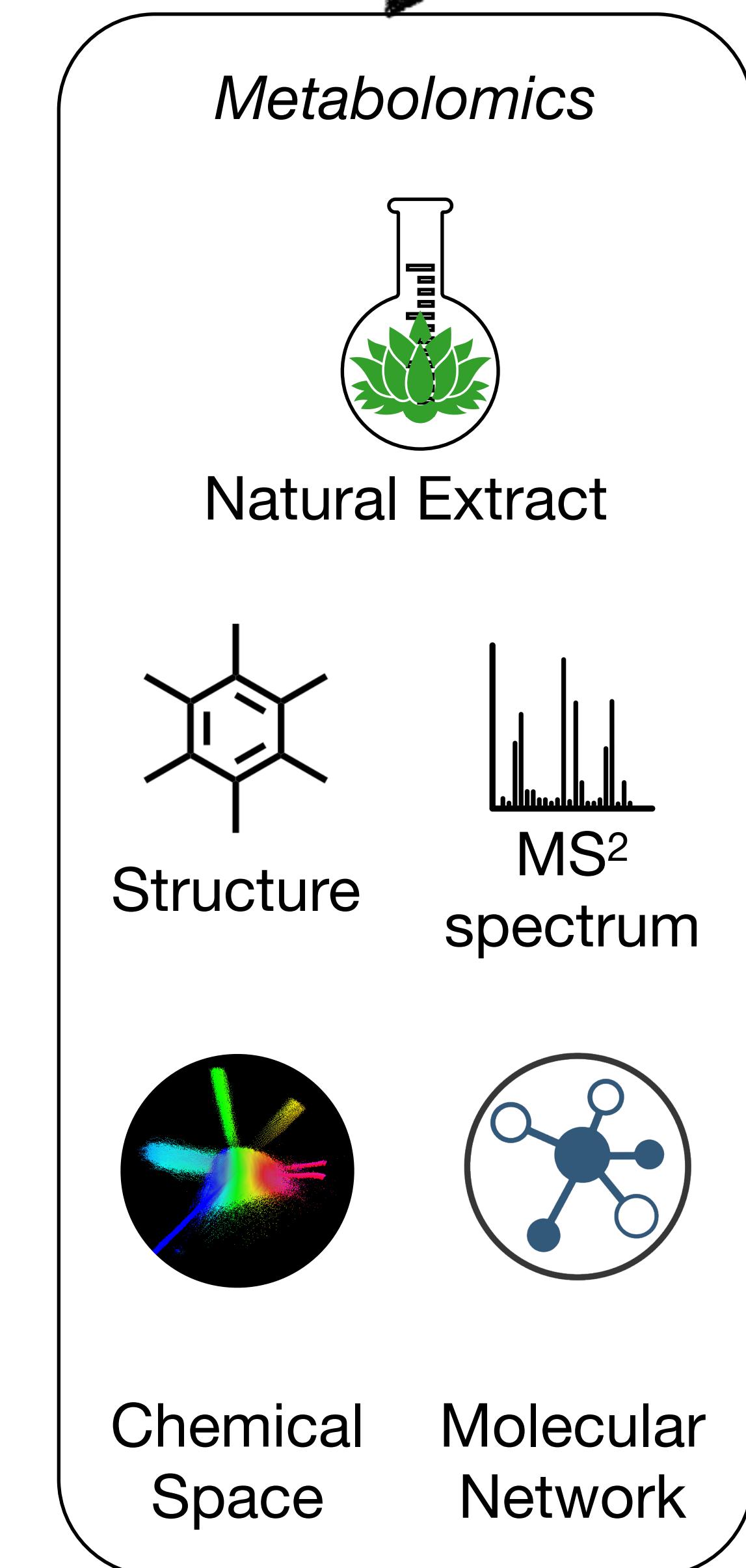
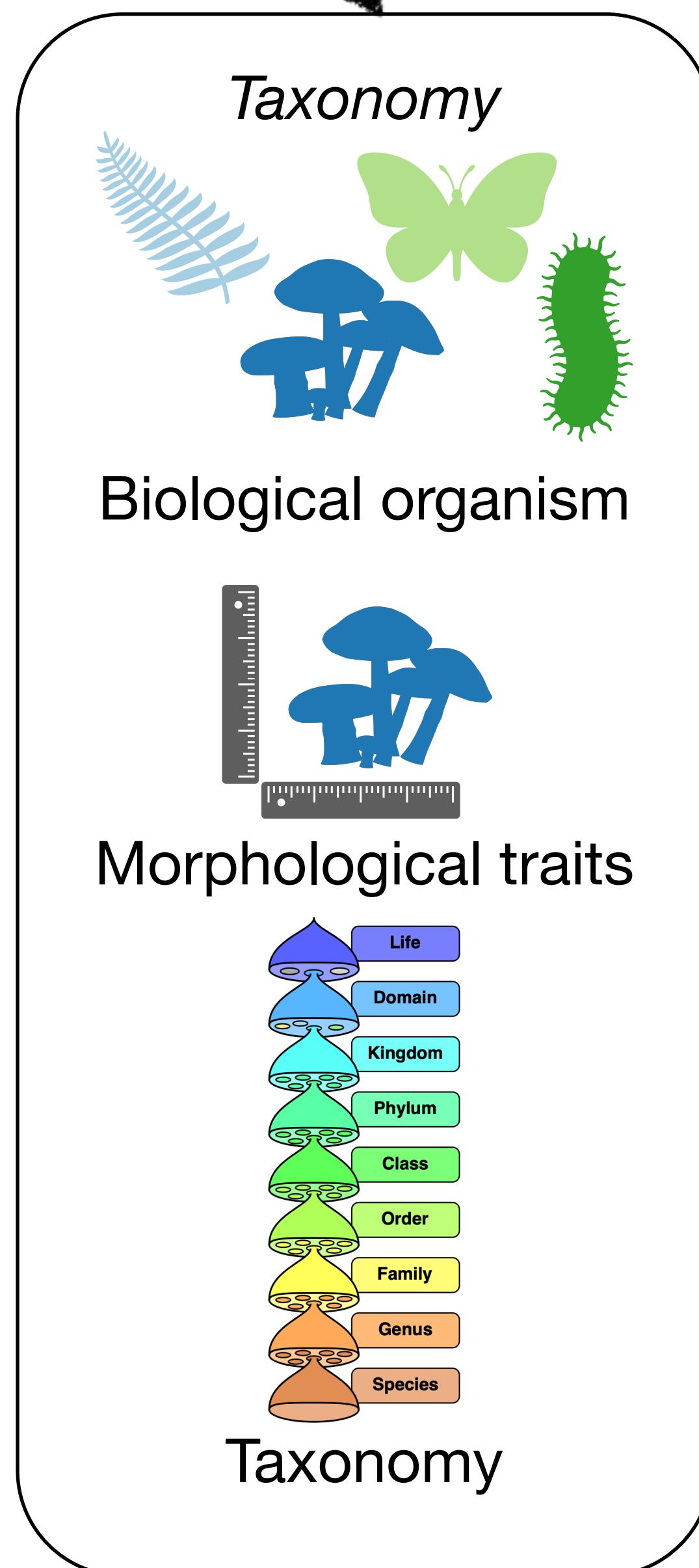
Taxonomically informed metabolite annotation



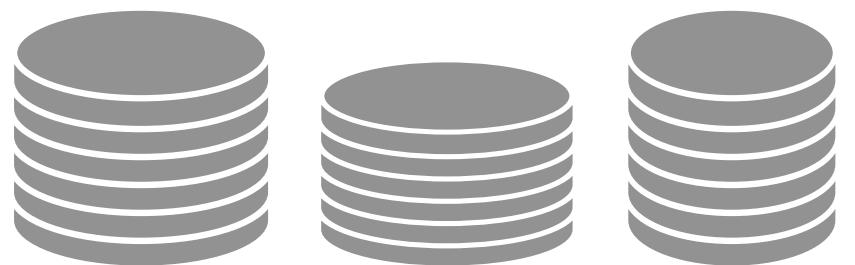
Taxonomically informed metabolite annotation



Taxonomically informed metabolite annotation



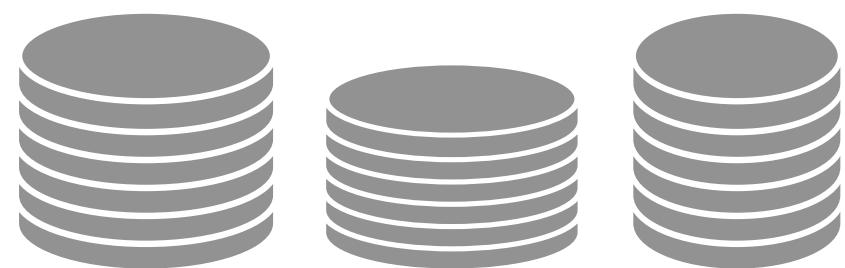
The initiative



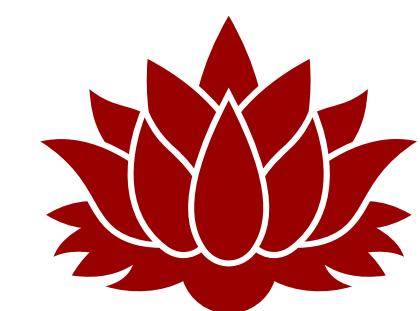
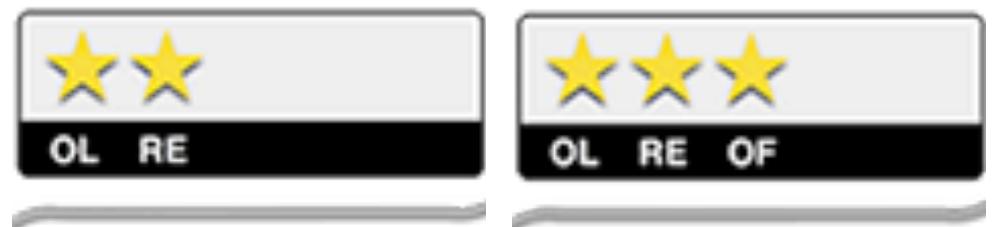
Available Natural
Product Data



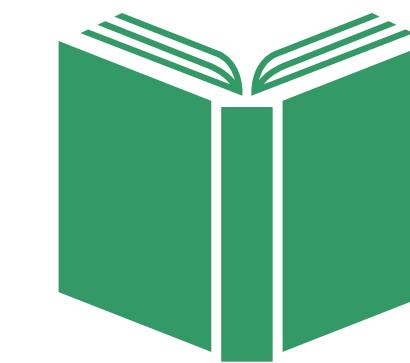
The initiative



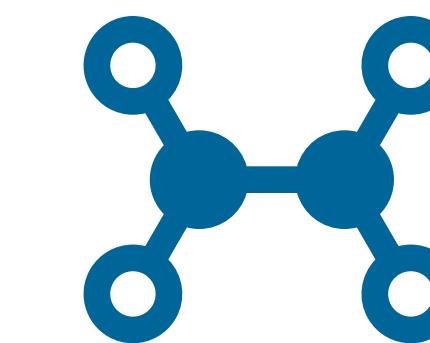
Available Natural
Product Data



Biological
Taxon



Reference

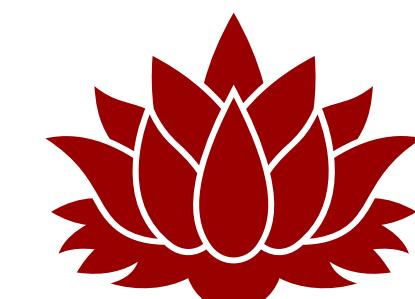


Chemical
Structure

The initiative



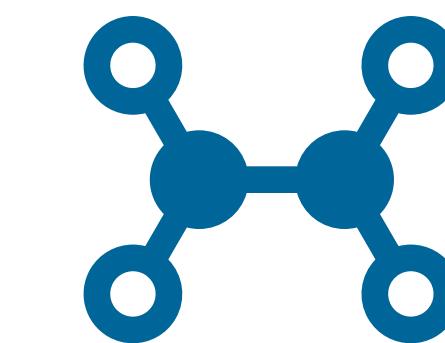
Available Natural
Product Data



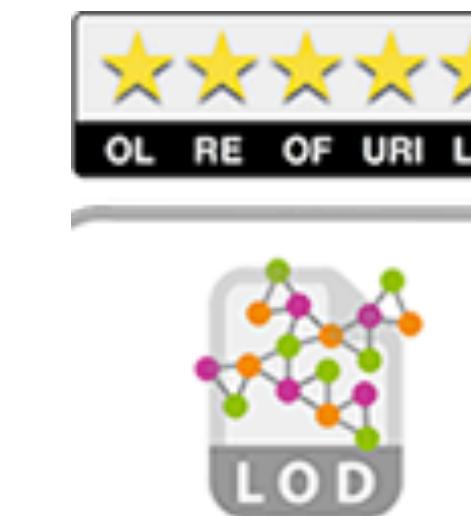
Biological
Taxon



Reference



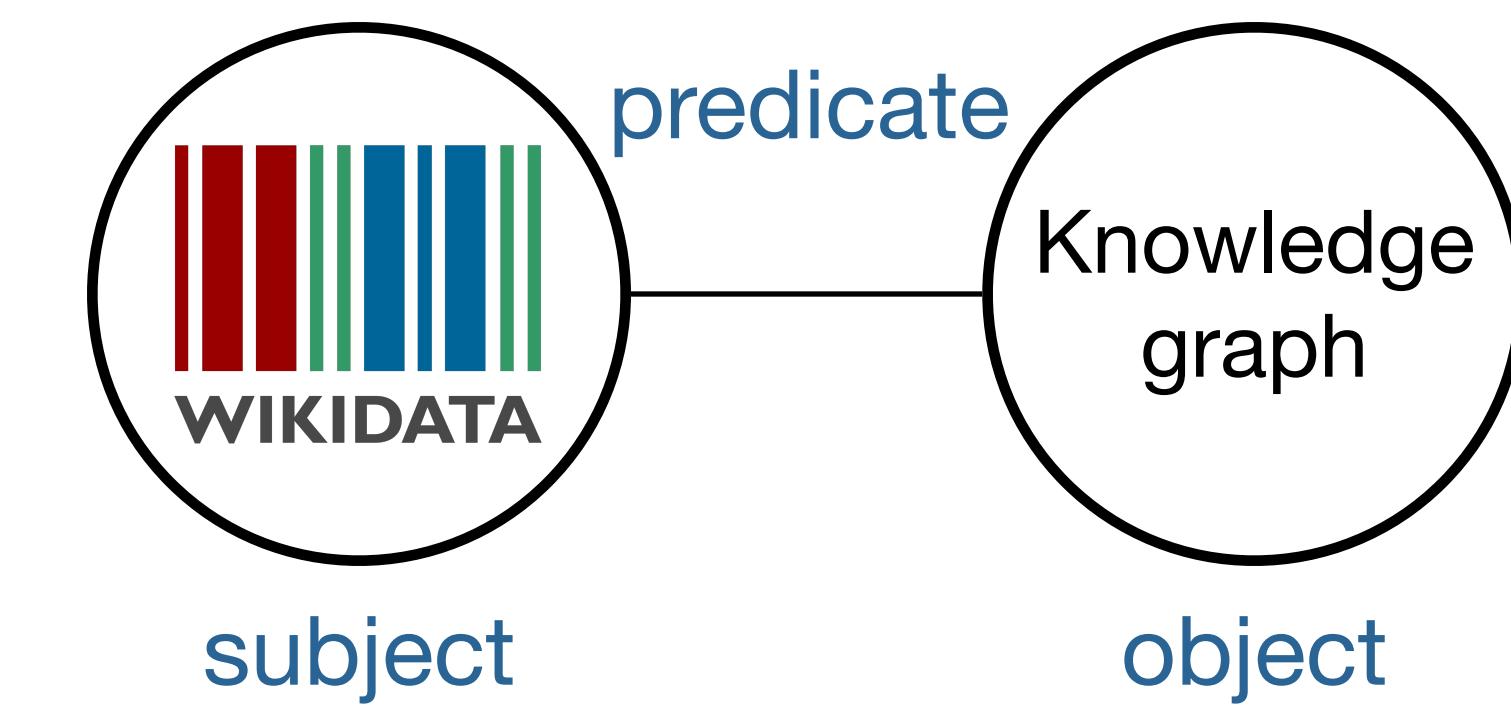
Chemical
Structure



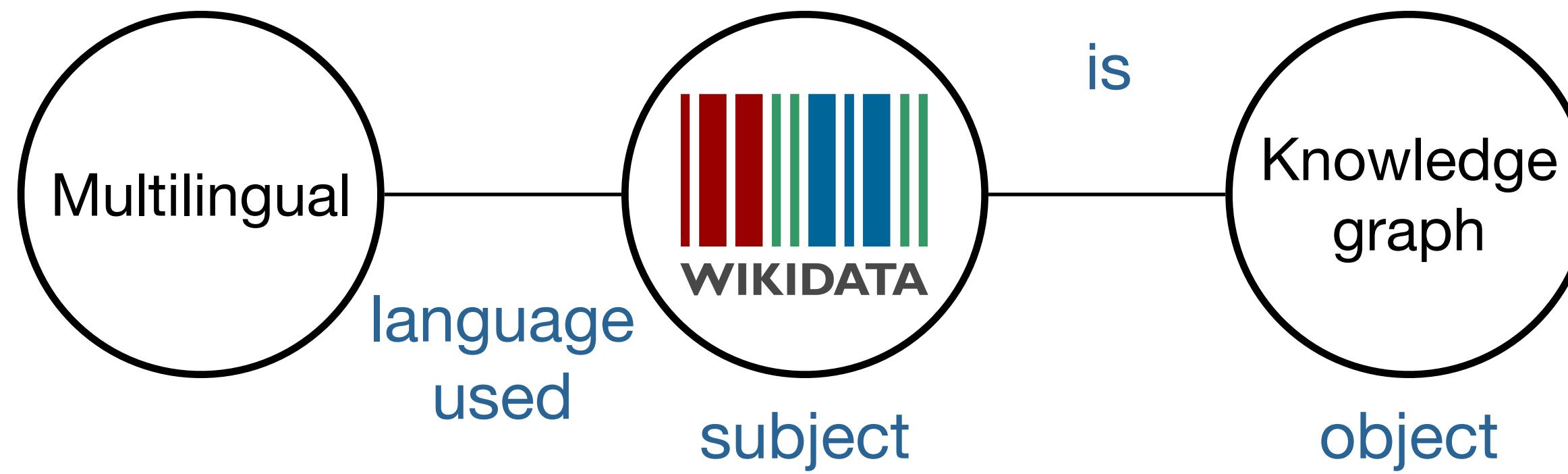
The initiative - Wikidata



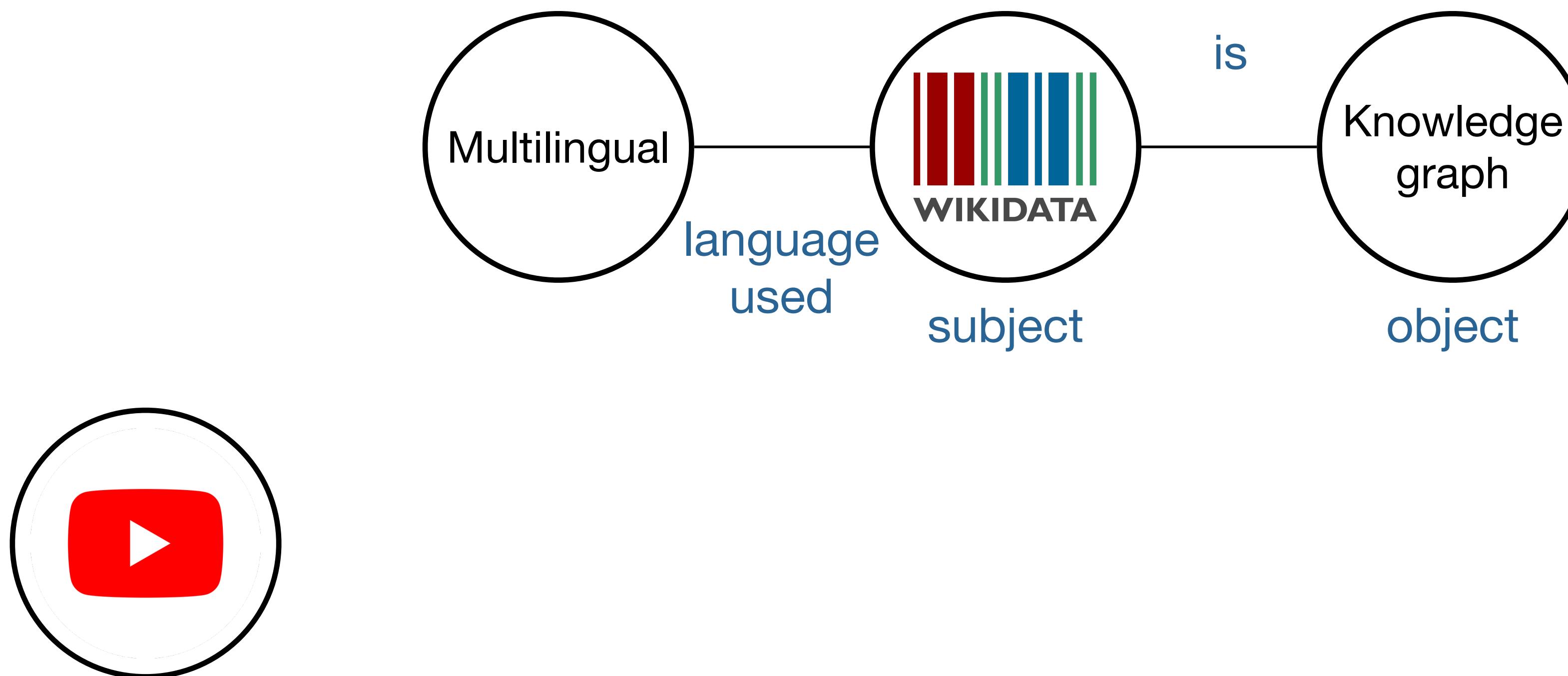
The initiative - Wikidata



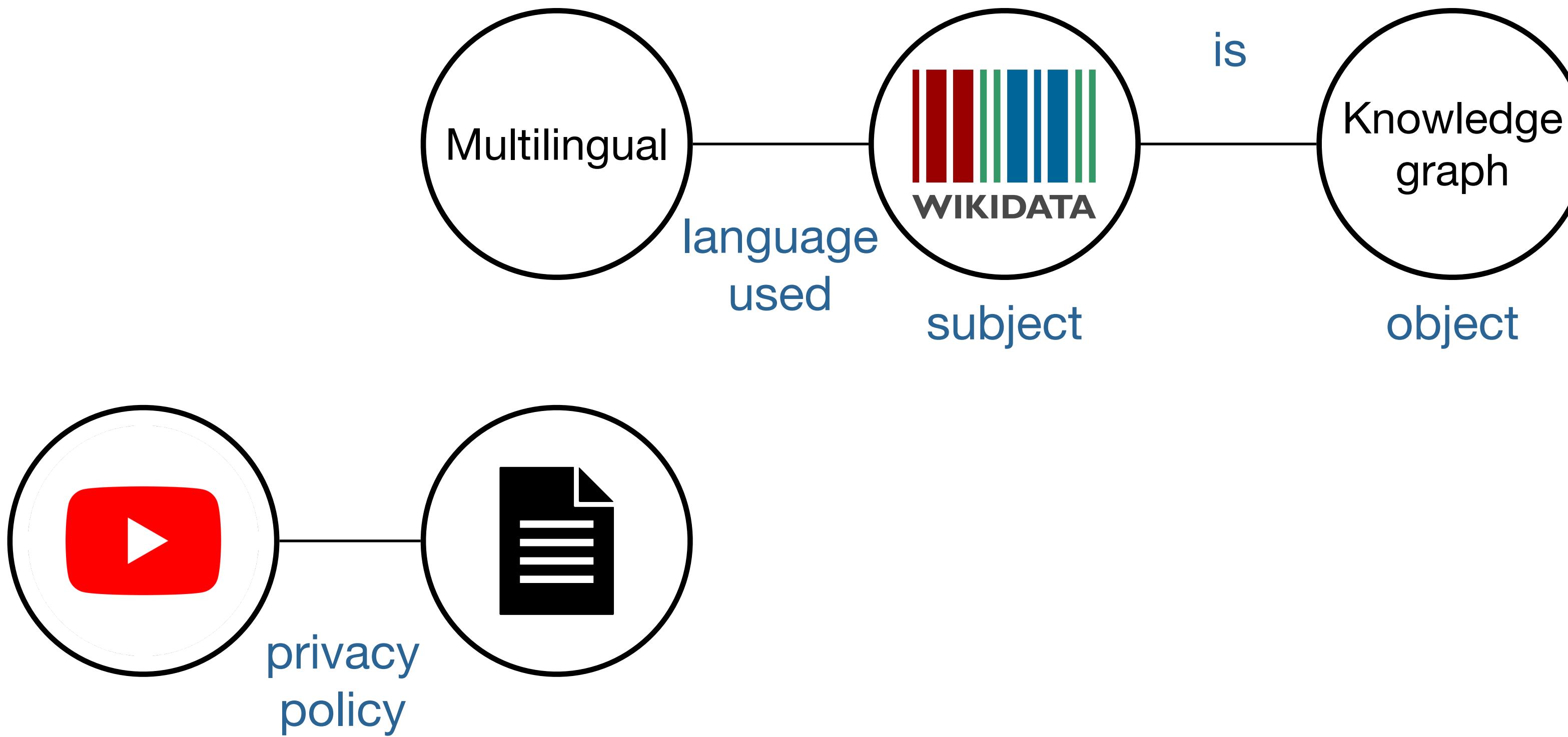
The initiative - Wikidata



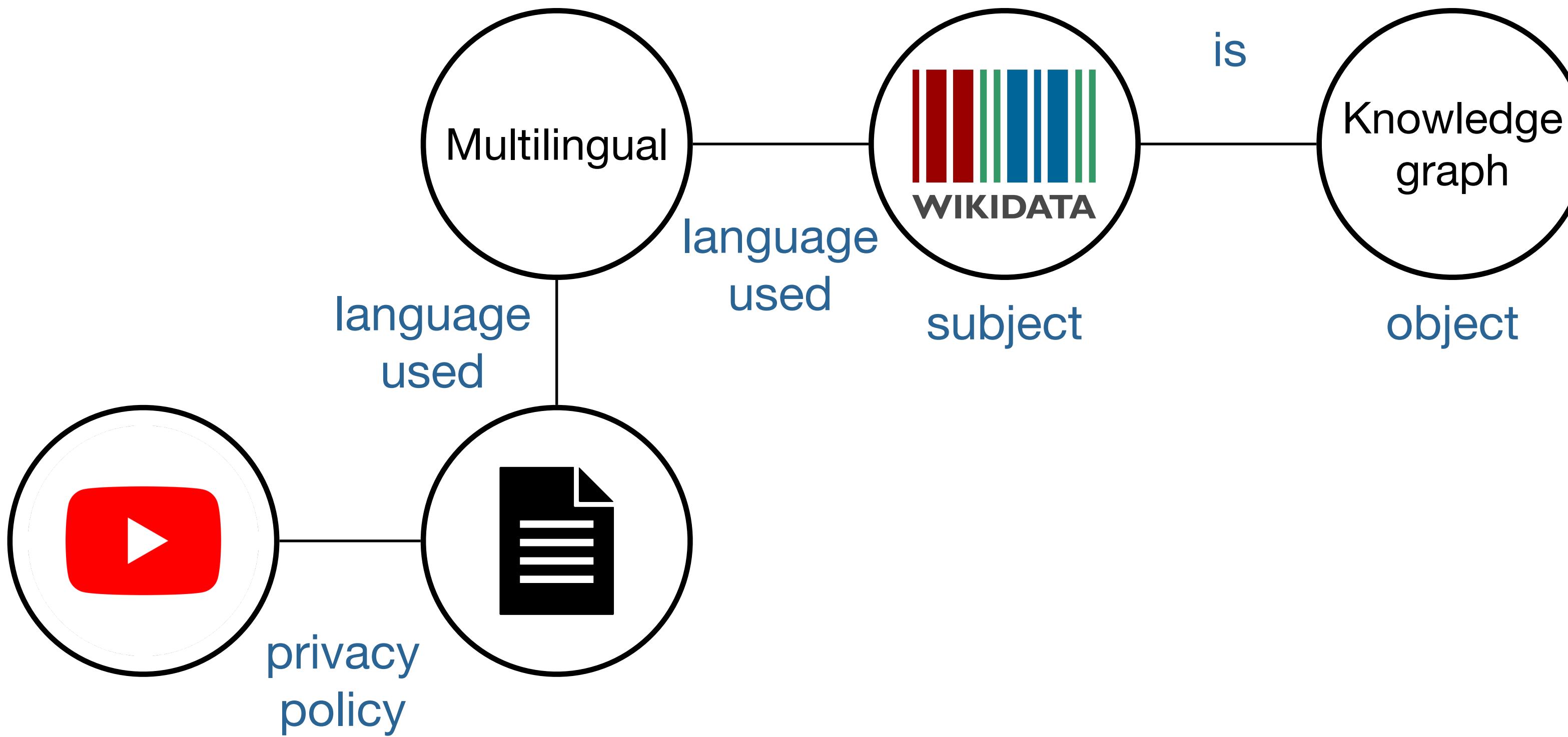
The initiative - Wikidata



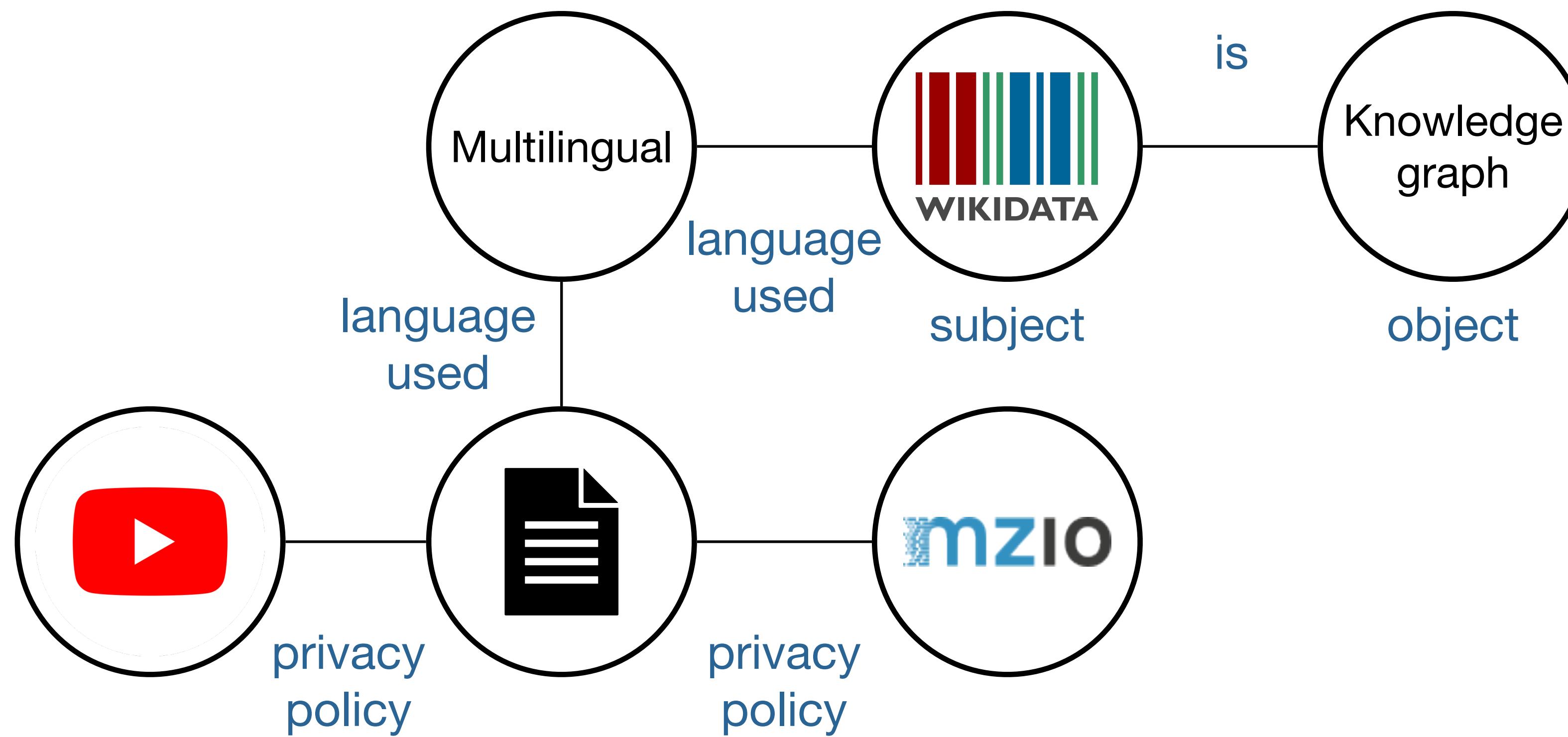
The initiative - Wikidata



The initiative - Wikidata

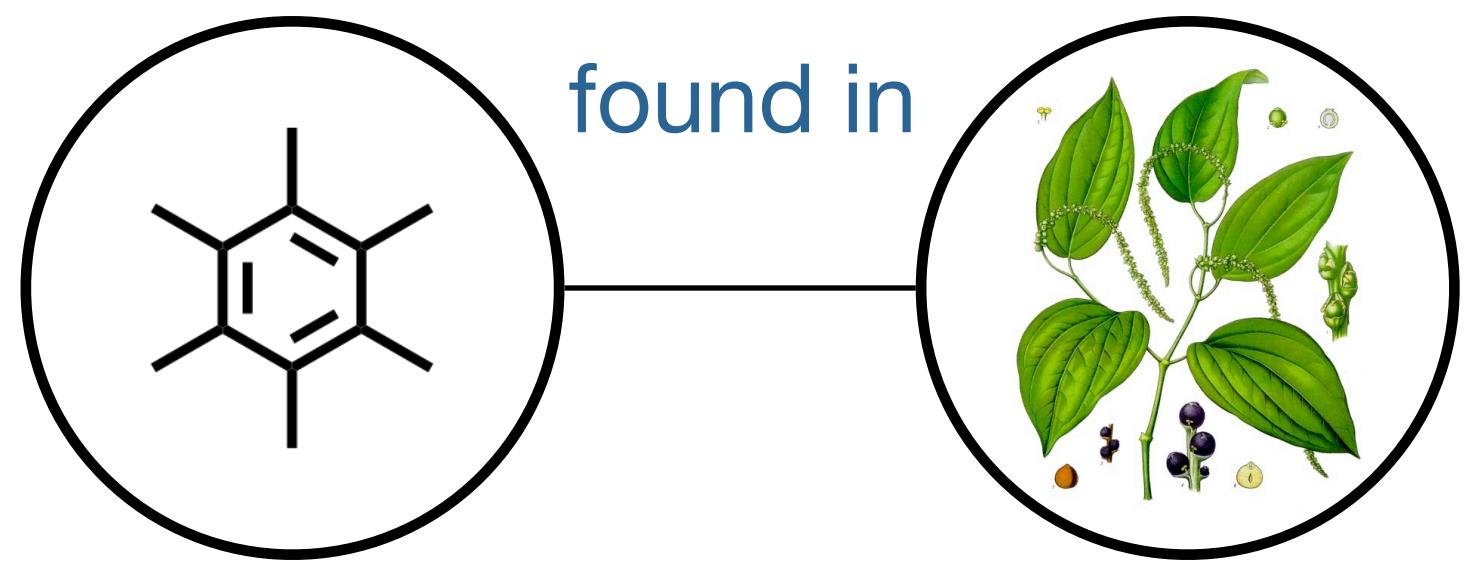


The initiative - Wikidata



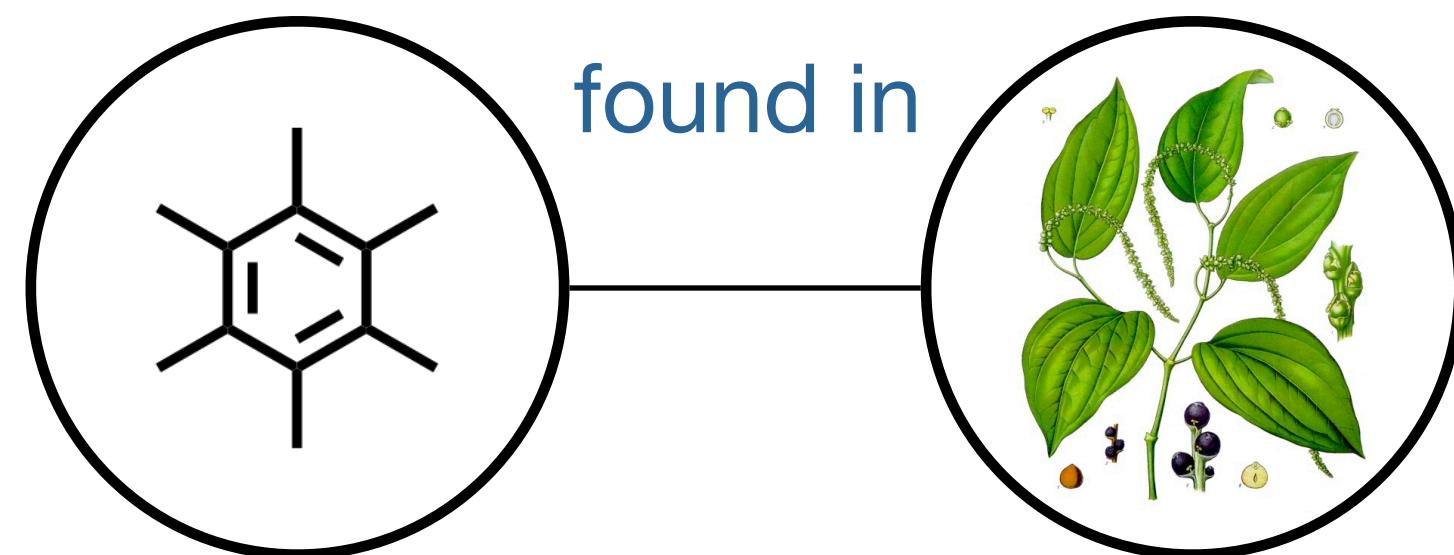
The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Piper nigrum* L.? »



The initiative - SPARQL

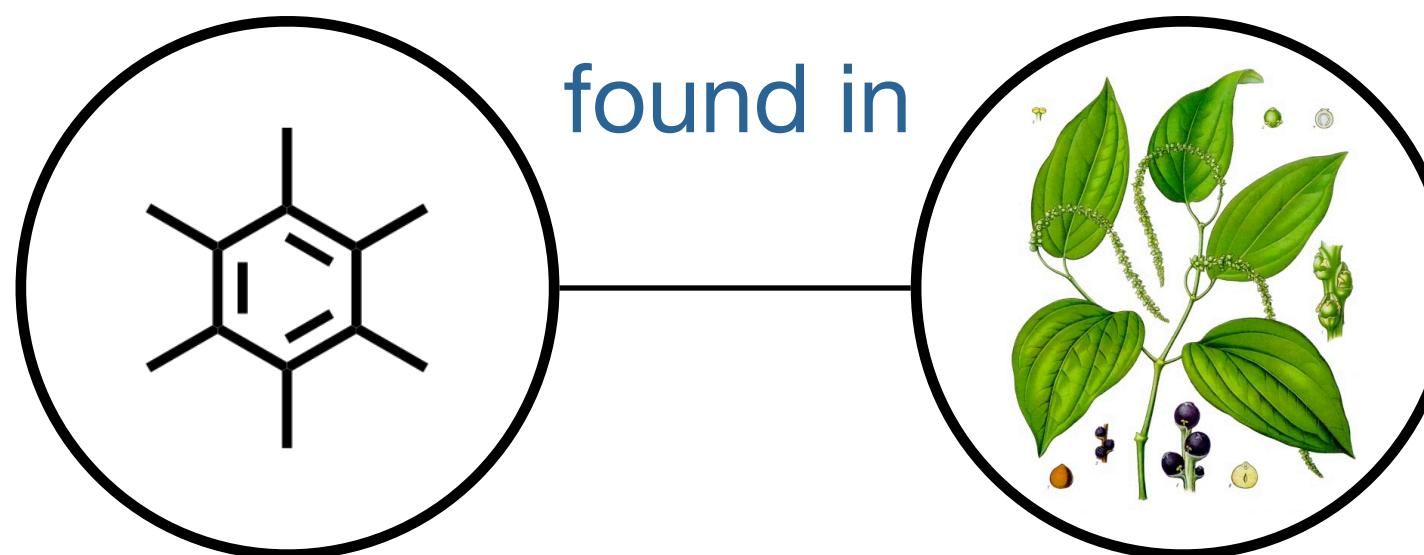
« Hey Wiki, what are the compounds found in *Piper nigrum* L.? »



```
SELECT DISTINCT  
?structure ?structure_smiles ?structure_inchikey  
WHERE {  
VALUES ?organism {  
wd:Q43084 # Piper nigrum  
}  
?organism_child (wdt:P171*) ?organism.  
?structure wdt:P233 ?structure_smiles;  
wdt:P235 ?structure_inchikey;  
(p:P703/ps:P703) ?organism_child.  
}
```

The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Piper nigrum* L.? »



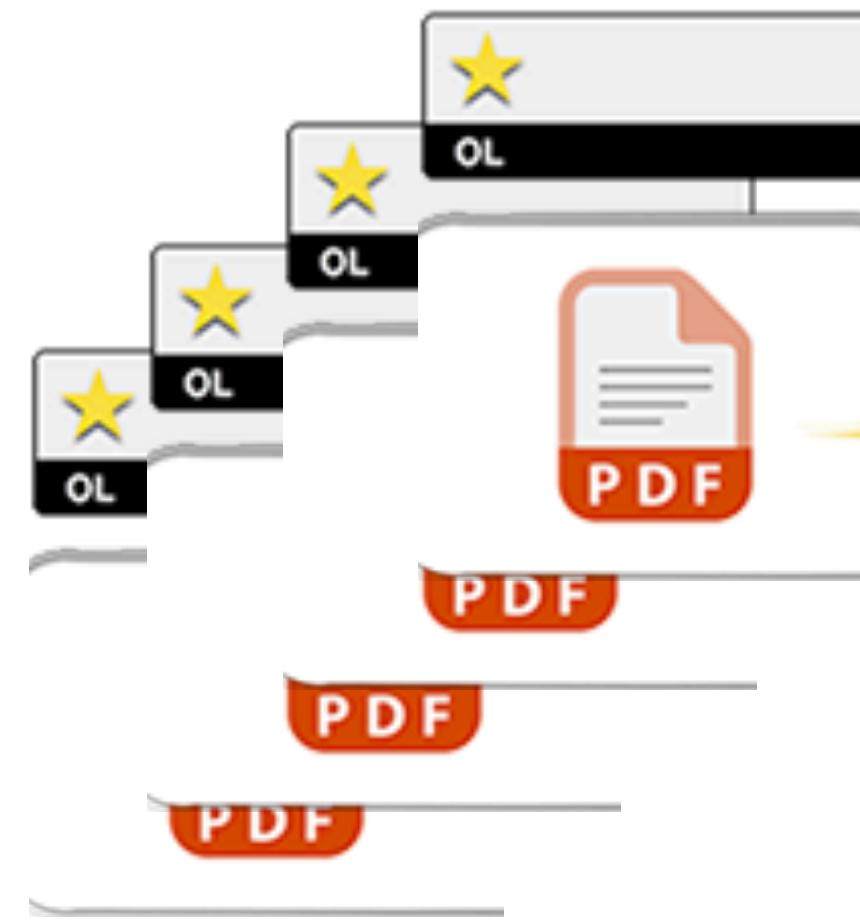
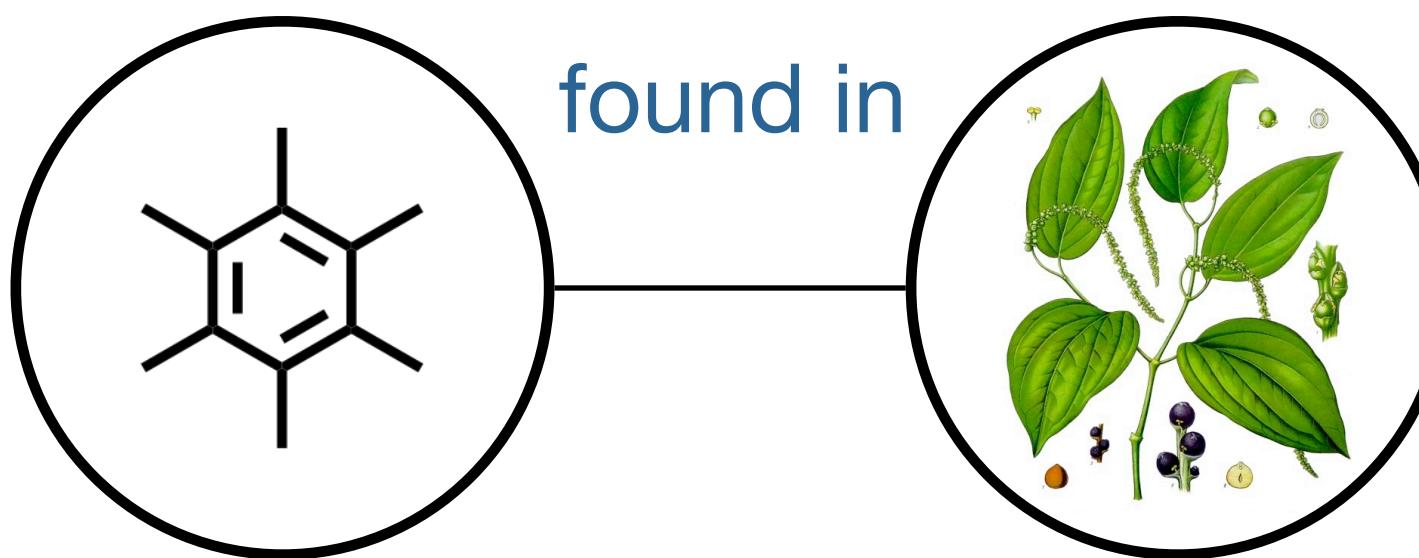
```
SELECT DISTINCT  
?structure ?structure_smiles ?structure_inchikey  
WHERE {  
VALUES ?organism {  
wd:Q43084 # Piper nigrum  
}  
?organism_child (wdt:P171*) ?organism.  
?structure wdt:P233 ?structure_smiles;  
wdt:P235 ?structure_inchikey;  
(p:P703/ps:P703) ?organism_child.  
}
```

row	structure	structure_smiles	structure_inchikey
1	wd:Q105143500		KNOXUMZPTHELAO-UHFFFAOYSA-N
...
419	wd:Q105224845		QOESZPJYTVEXNN-UHFFFAOYSA-N

419 results in 0.5 second

The initiative - SPARQL

« Hey Wiki, what are the compounds found in *Piper nigrum* L.? »



```
SELECT DISTINCT  
?structure ?structure_smiles ?structure_inchikey  
WHERE {  
VALUES ?organism {  
wd:Q43084 # Piper nigrum  
}  
?organism_child (wdt:P171*) ?organism.  
?structure wdt:P233 ?structure_smiles;  
wdt:P235 ?structure_inchikey;  
(p:P703/ps:P703) ?organism_child.  
}
```

row	structure	structure_smiles	structure_inchikey
1	wd:Q105143500		KNOXUMZPTHELAO-UHFFFAOYSA-N
...
419	wd:Q105224845		QOESZPJYTVEXNN-UHFFFAOYSA-N

419 results in 0.5 second

The initiative - SPARQL

« Hey Wiki, what are the compounds found in Ranunculaceae? »

3,398 results in 0.9 second

The initiative - SPARQL

« Hey Wiki, what are the compounds found in Ranunculaceae? »

3,398 results in 0.9 second

« Hey Wiki, what are the compounds found in Ranunculales? »

6,350 results in 1.4 seconds

The initiative - SPARQL

« Hey Wiki, what are the compounds found in Ranunculaceae? »

3,398 results in 0.9 second

« Hey Wiki, what are the compounds found in Ranunculales? »

6,350 results in 1.4 seconds

« Sorry Wiki, I forgot, how many species belong to Ranunculales? »

11,669 results in 1.1 seconds

The initiative - SPARQL

« Hey Wiki, what are the compounds found in Ranunculaceae? »

3,398 results in 0.9 second

« Sorry Wiki, I forgot, how many species belong to Ranunculales? »

11,669 results in 1.1 seconds

« Hey Wiki, what are the compounds found in Ranunculales? »

6,350 results in 1.4 seconds

« Ok Wiki, among the compounds found in the Ranunculales how many were already described as bitter? »

59 results in 1.4 seconds



The initiative - SPARQL

« Ok Wiki, which genera contain compounds structurally similar to the ones reported as pungent in black pepper, and how many? List the top 5 only »

The initiative - SPARQL

« Ok Wiki, which genera contain compounds structurally similar to the ones reported as pungent in black pepper, and how many? List the top 5 only »

5 results in 1.1 seconds

row	taxon	taxon_name	count
1	wd:Q159536	<i>Piper</i>	35
2	wd:Q163563	<i>Zingiber</i>	31
3	wd:Q133290	<i>Syzygium</i>	4
4	wd:Q2607361	<i>Aframomum</i>	4
5	wd:Q27583	<i>Bellis</i>	3



Developed right here!

For more: <https://doi.org/10.1186/s13321-018-0282-y>



The initiative - Community curation

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

hyperelodione D (Q116482353)

chemical compound edit

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

Statements

instance of	chemical compound 0 references
-------------	-----------------------------------

mass	548.797 dalton 1 reference based on heuristic inferred from SMILES
------	--

chemical formula	C ₃₆ H ₅₂ O ₄ 1 reference based on heuristic inferred from SMILES
------------------	--

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes
Random Lexeme

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

The initiative - Community curation

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

hyperelodione D (Q116482353)

chemical compound edit

Language	Label	Description	Also known as
English	hyperelodione D	chemical compound	

canonical SMILES

C/C(C)=C\CC/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(C)(C)O2

reason for deprecated rank structural reassignment
1 reference

stated in Cytotoxic polyprenylated phloroglucinol derivatives from Hypericum elodeoides Choisy modulating the transactivation of RXRa

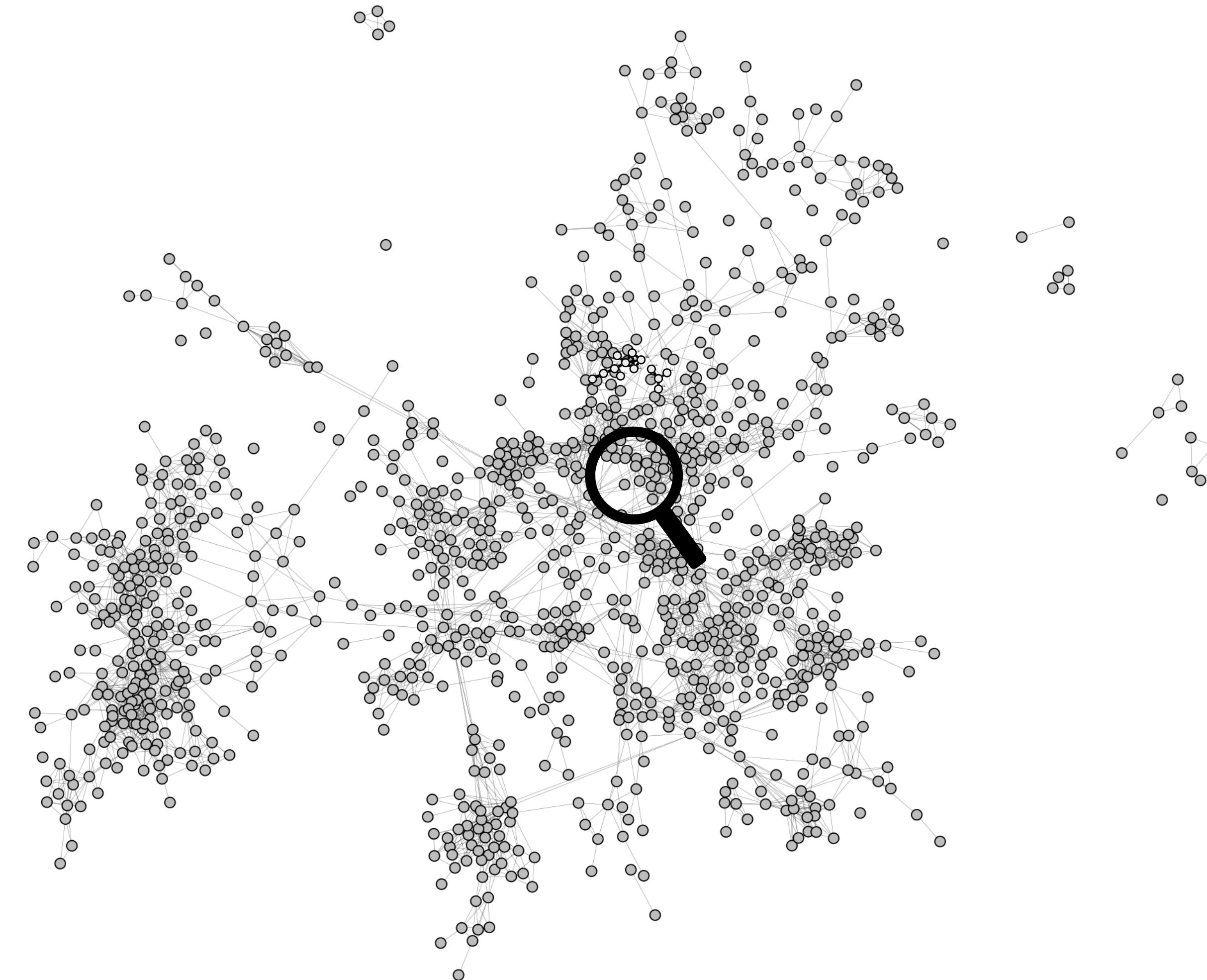
C/C(C)=C/CC1=C2C3(O[H])C(C4(C/C=C(C)/CC/C=C(C)/C)C1=O)(O[H])C(C(C)=CC4)([H])CC3([H])C(CC/C=C(C)/C)(C)O2

reason for preferred rank structural reassignment
1 reference

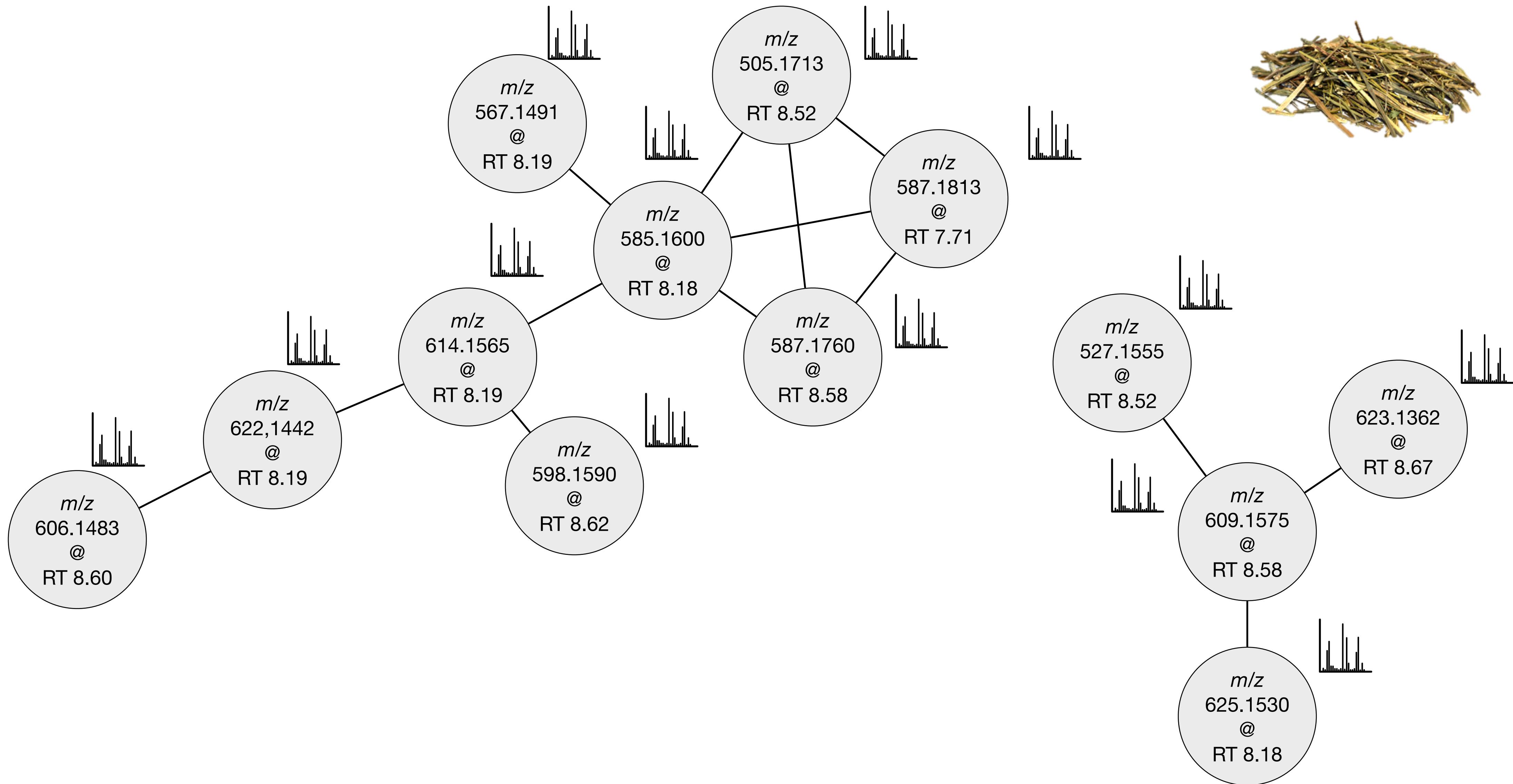
stated in Bioinspired Total Synthesis of Erectones A and B, and the Revised Structure of Hyperelodione D

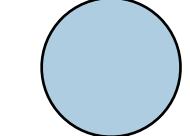
The initiative - back to annotations

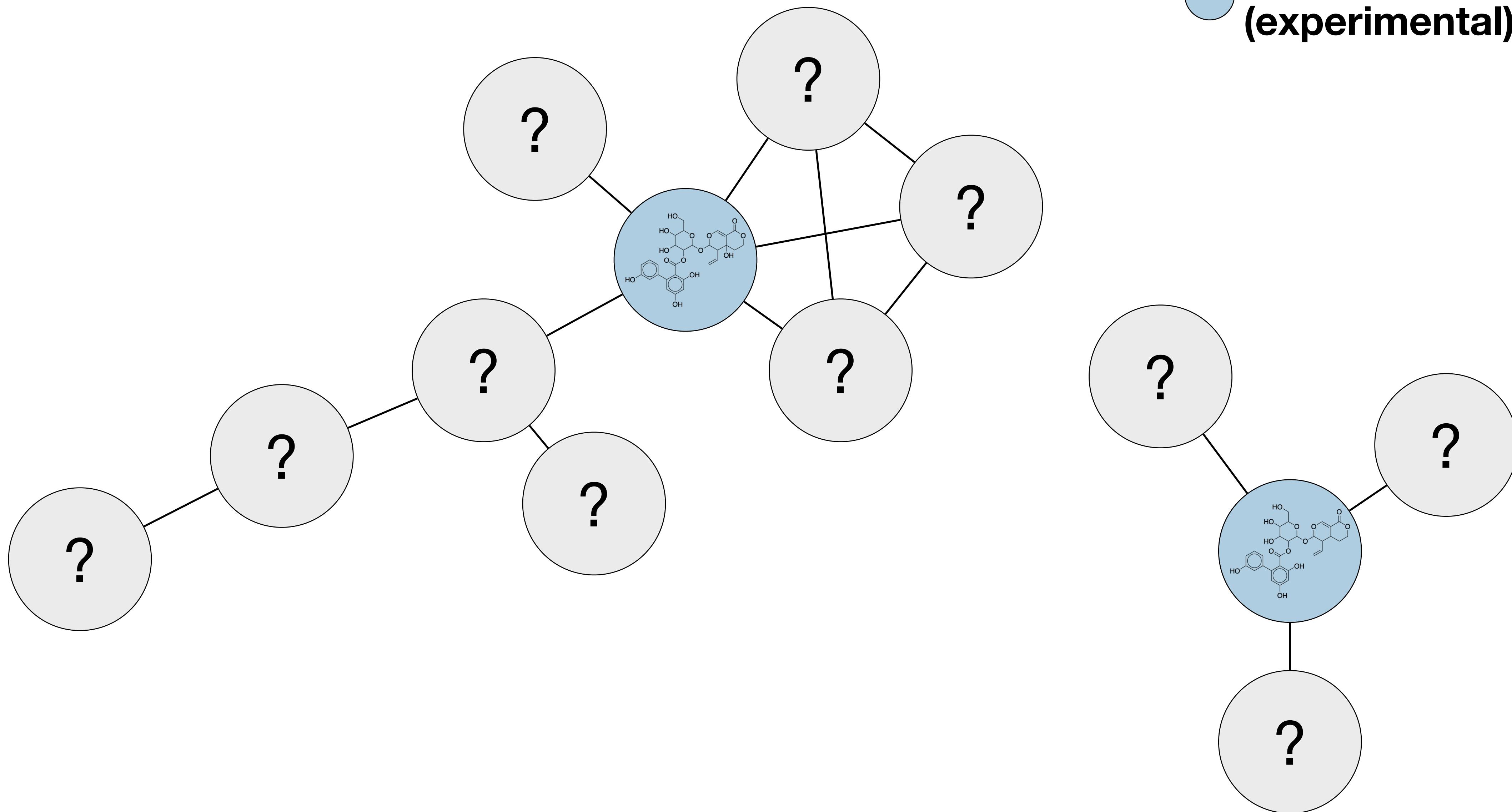
« But Wiki, can this help me with my annotations? »



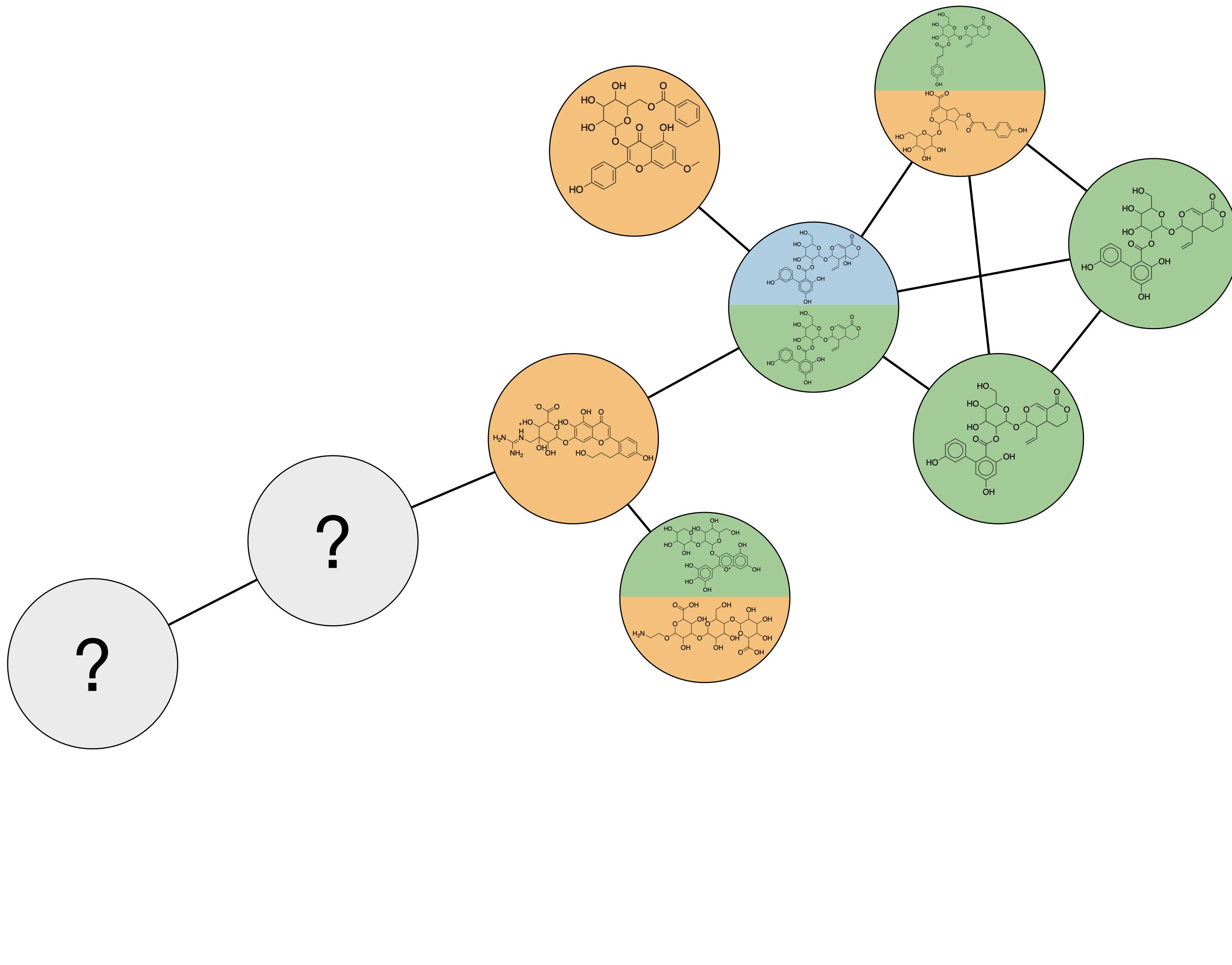
The initiative - back to annotations



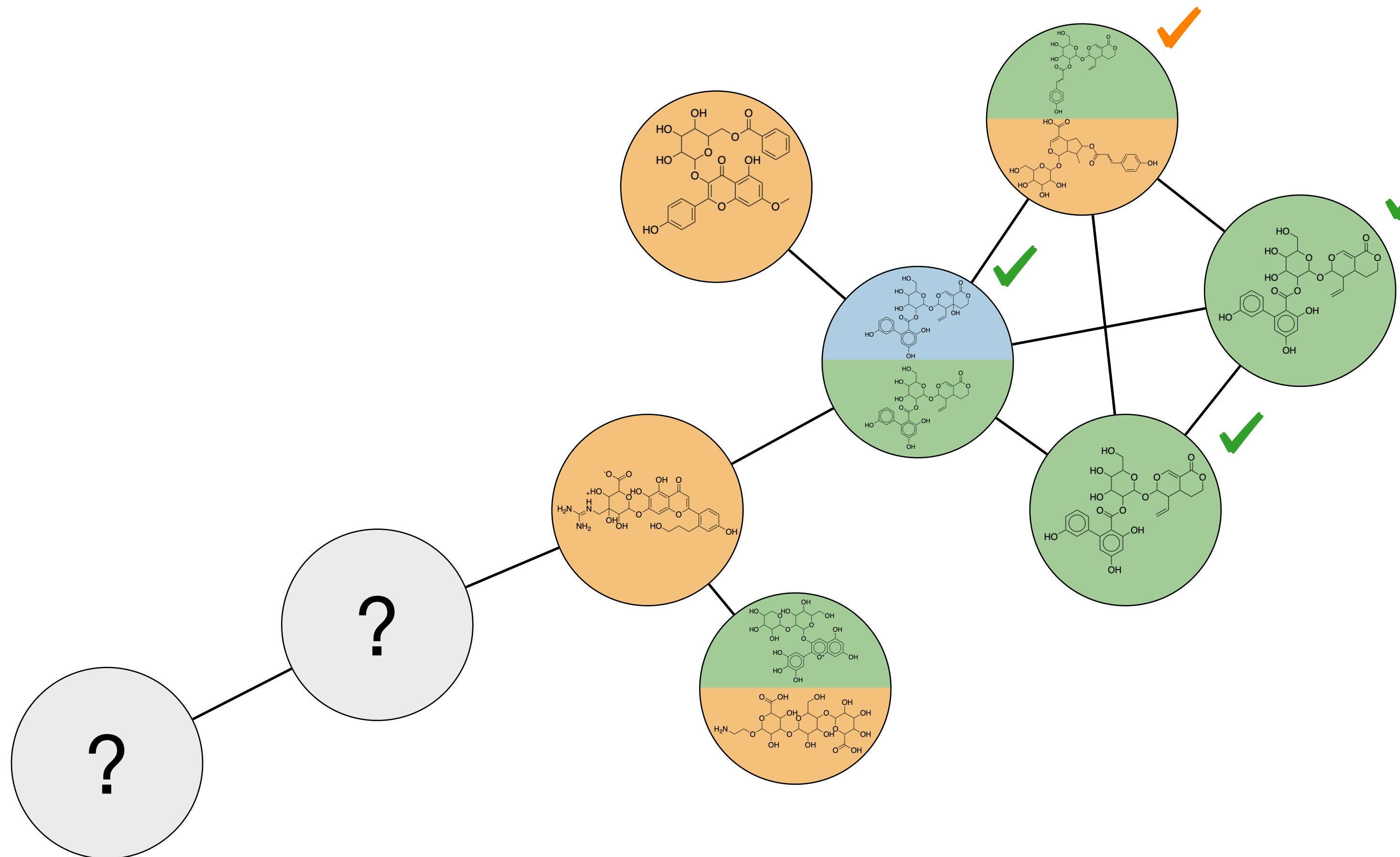
 **GNPS**
(experimental)



GNPS
(experimental)
ISDB-DNP (in silico)
Sirius (in silico)

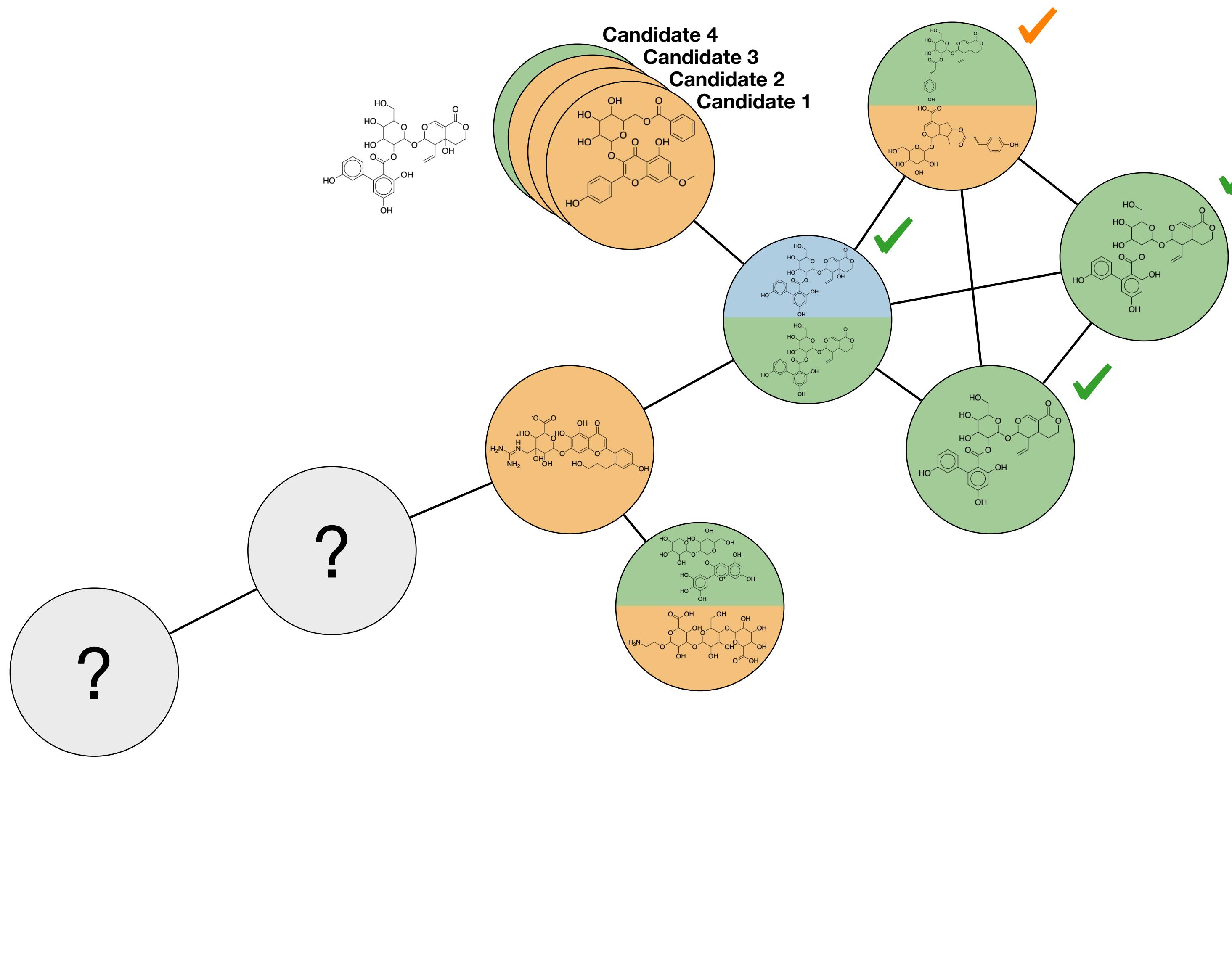


LOD-based annotation



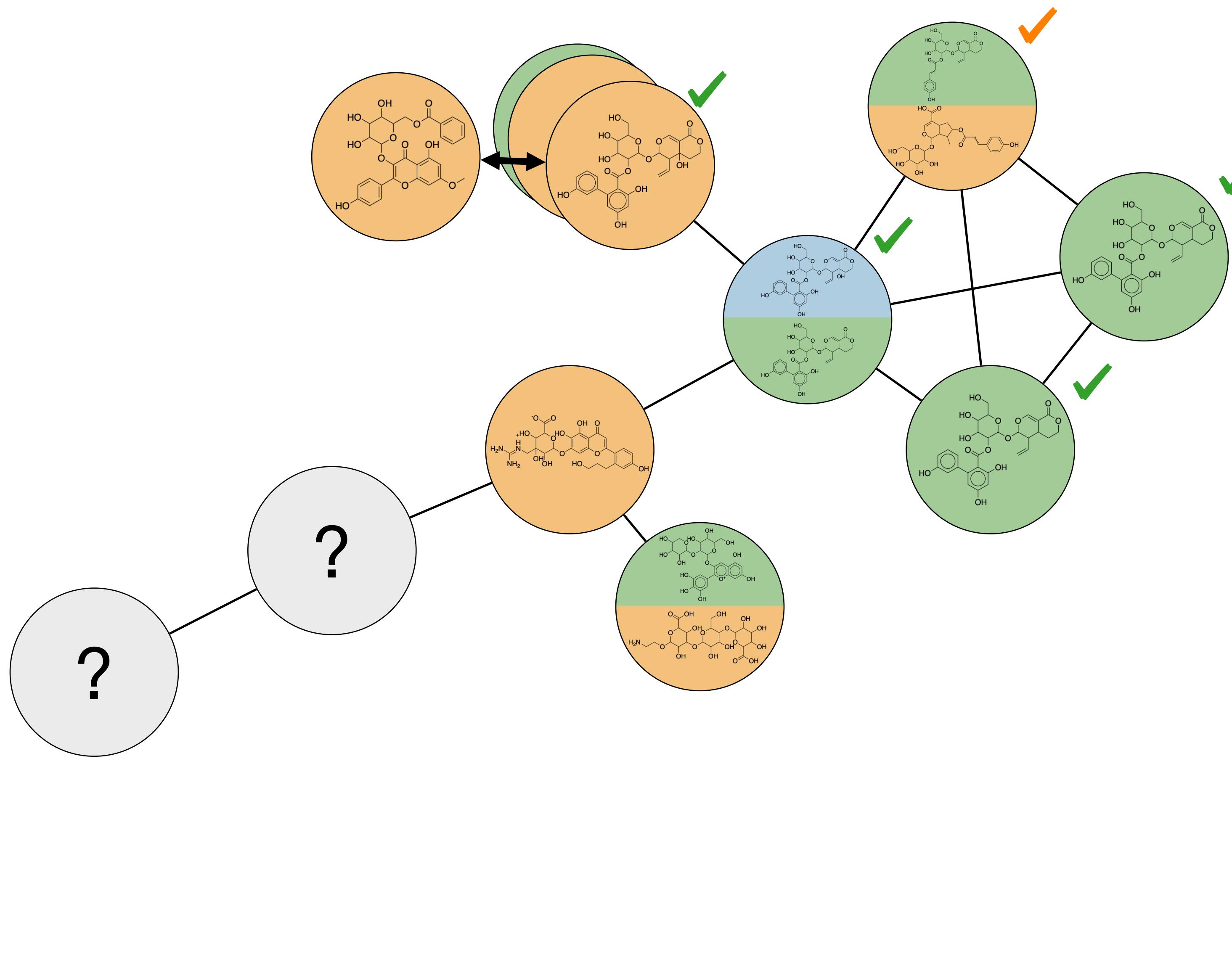
- GNPS (experimental)
- ISDB-DNP (in silico)
- Sirius (in silico)
- ✓ Found in the genus
- ✓ Found in the family

LOD-based annotation



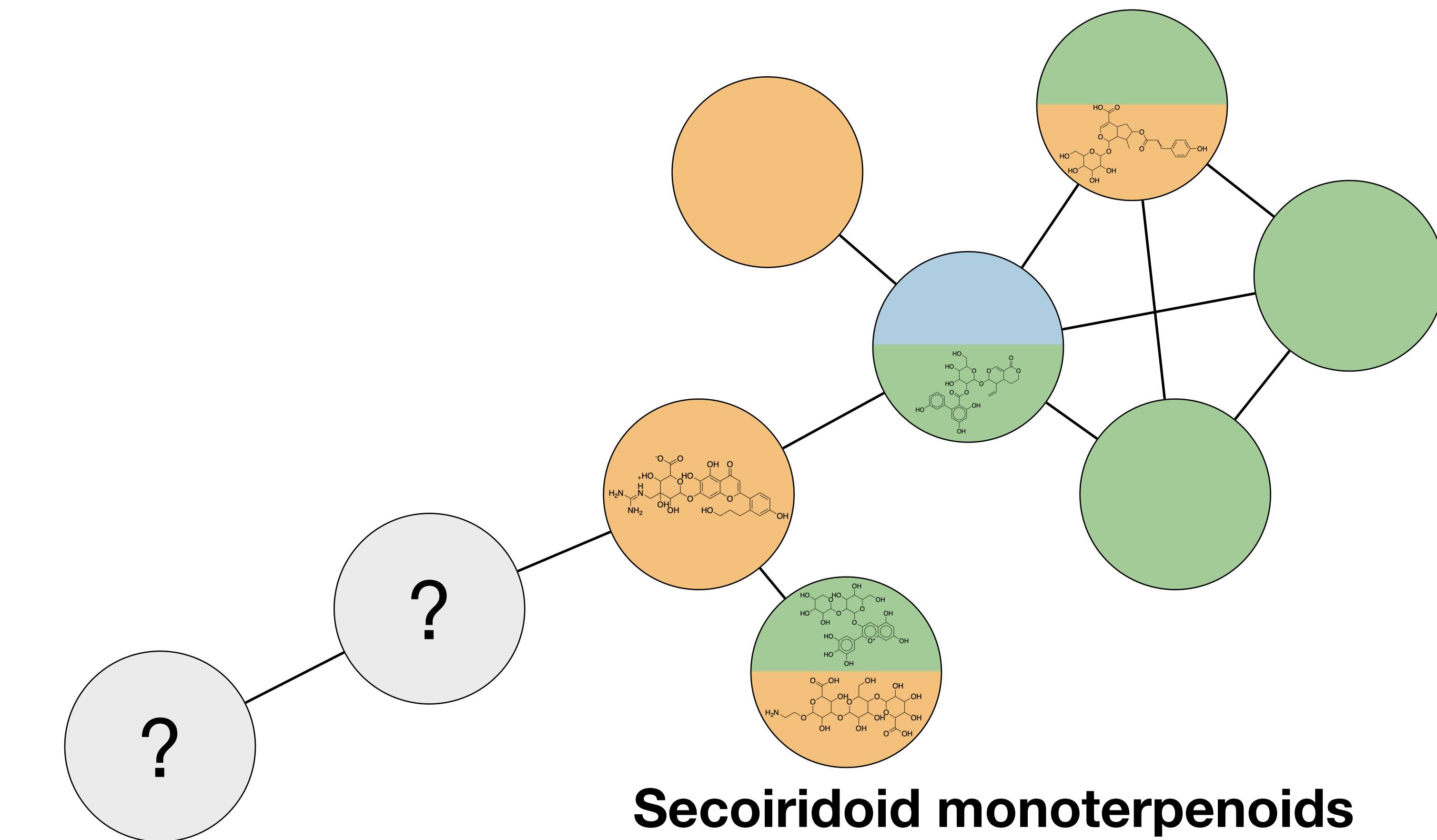
- GNPS (experimental)
- ISDB-DNP (in silico)
- Sirius (in silico)
- ✓ Found in the genus
- ✓ Found in the family

LOD-based annotation

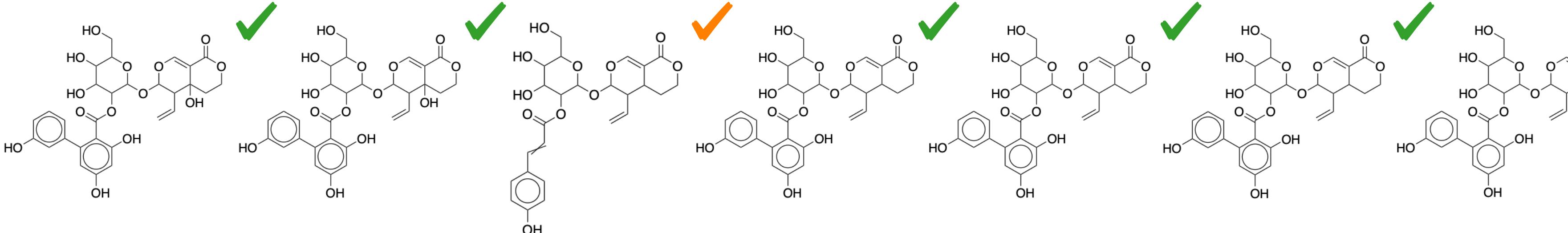


- GNPS (experimental)
- ISDB-DNP (in silico)
- Sirius (in silico)
- ✓ Found in the genus
- ✓ Found in the family

LOD-based annotation

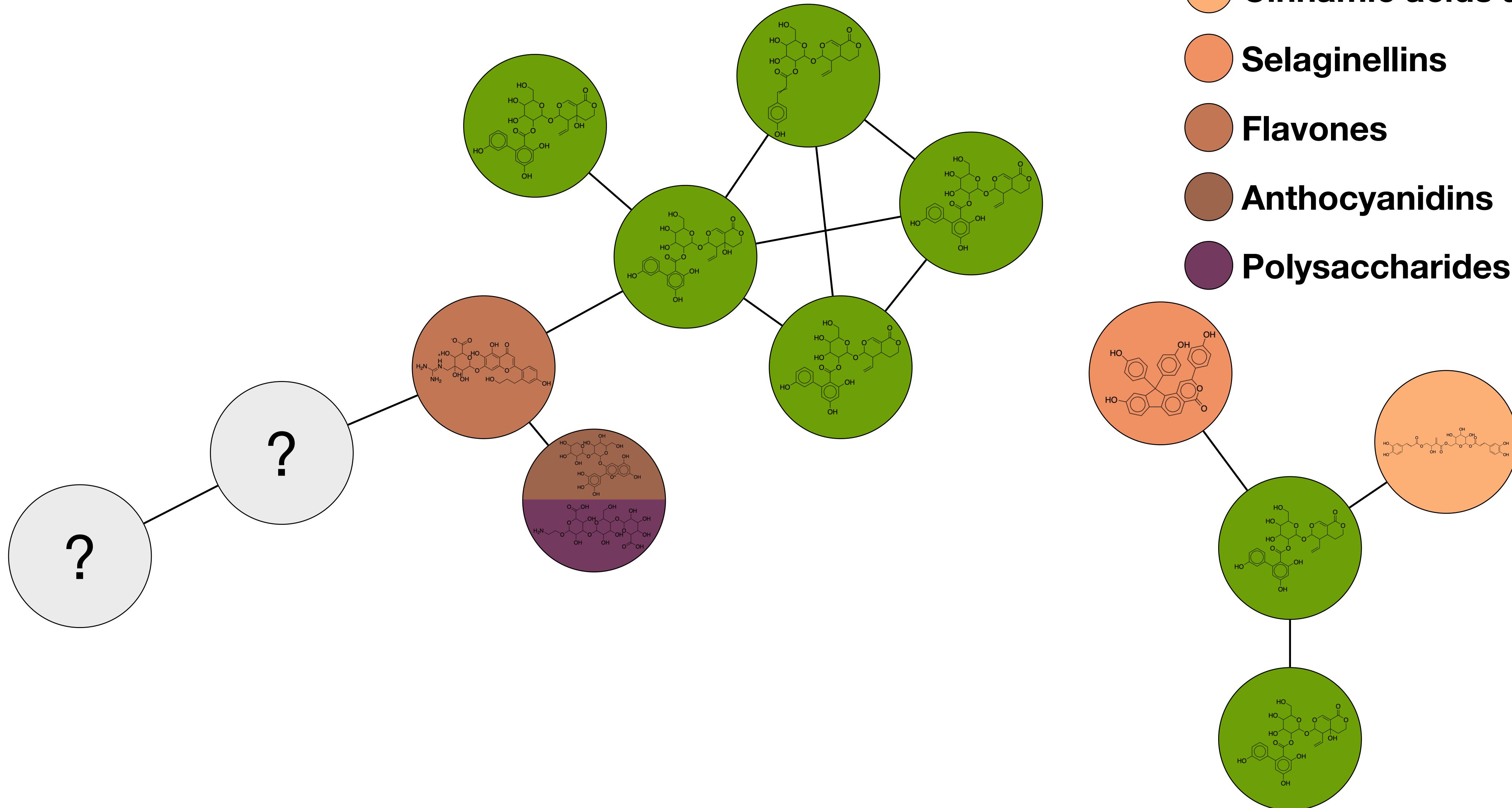


Secoiridoid monoterpenoids

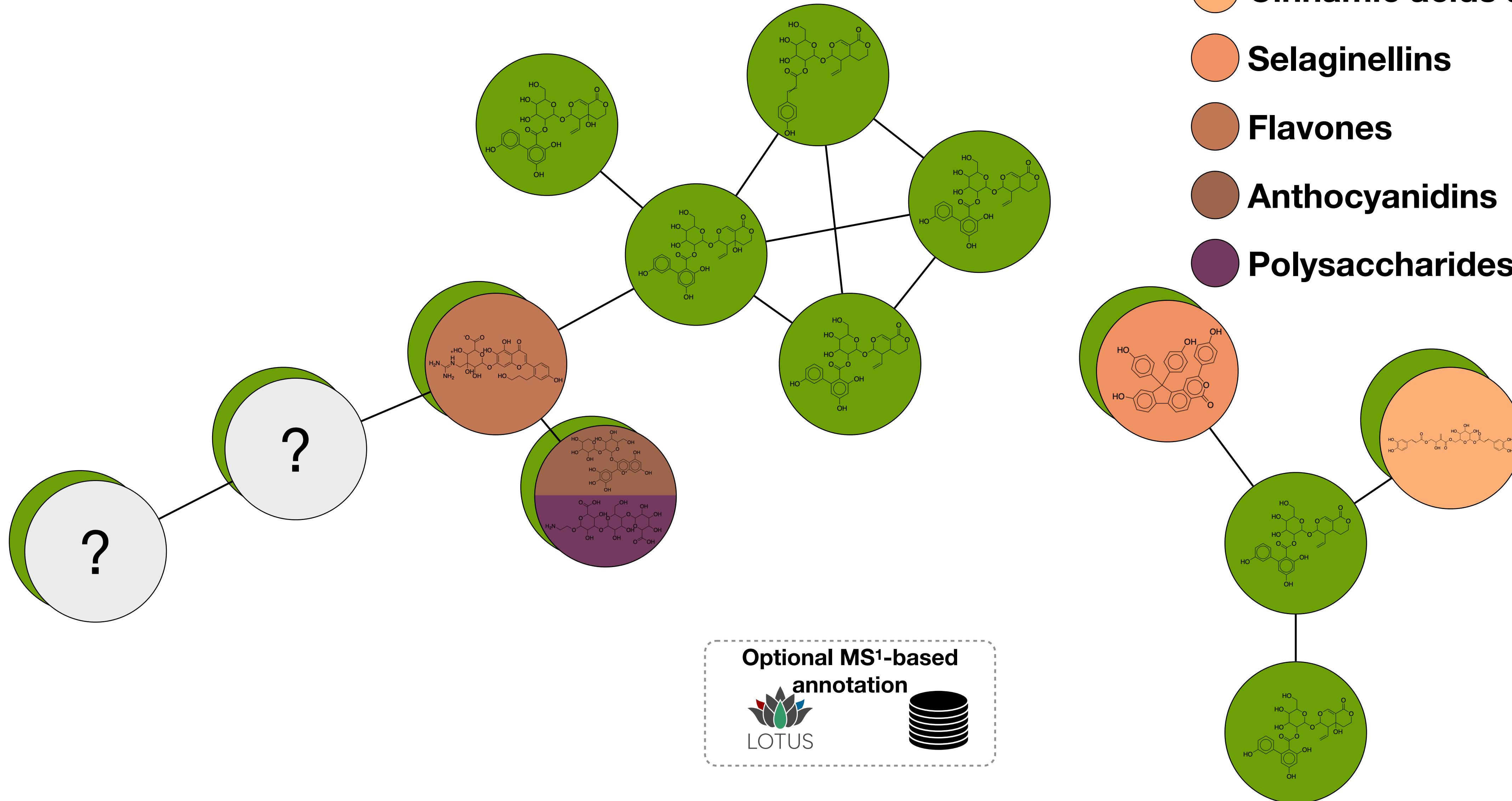


- GNPS (experimental)
- ISDB-DNP (in silico)
- Sirius (in silico)
- ✓ Found in the genus
- ✓ Found in the family

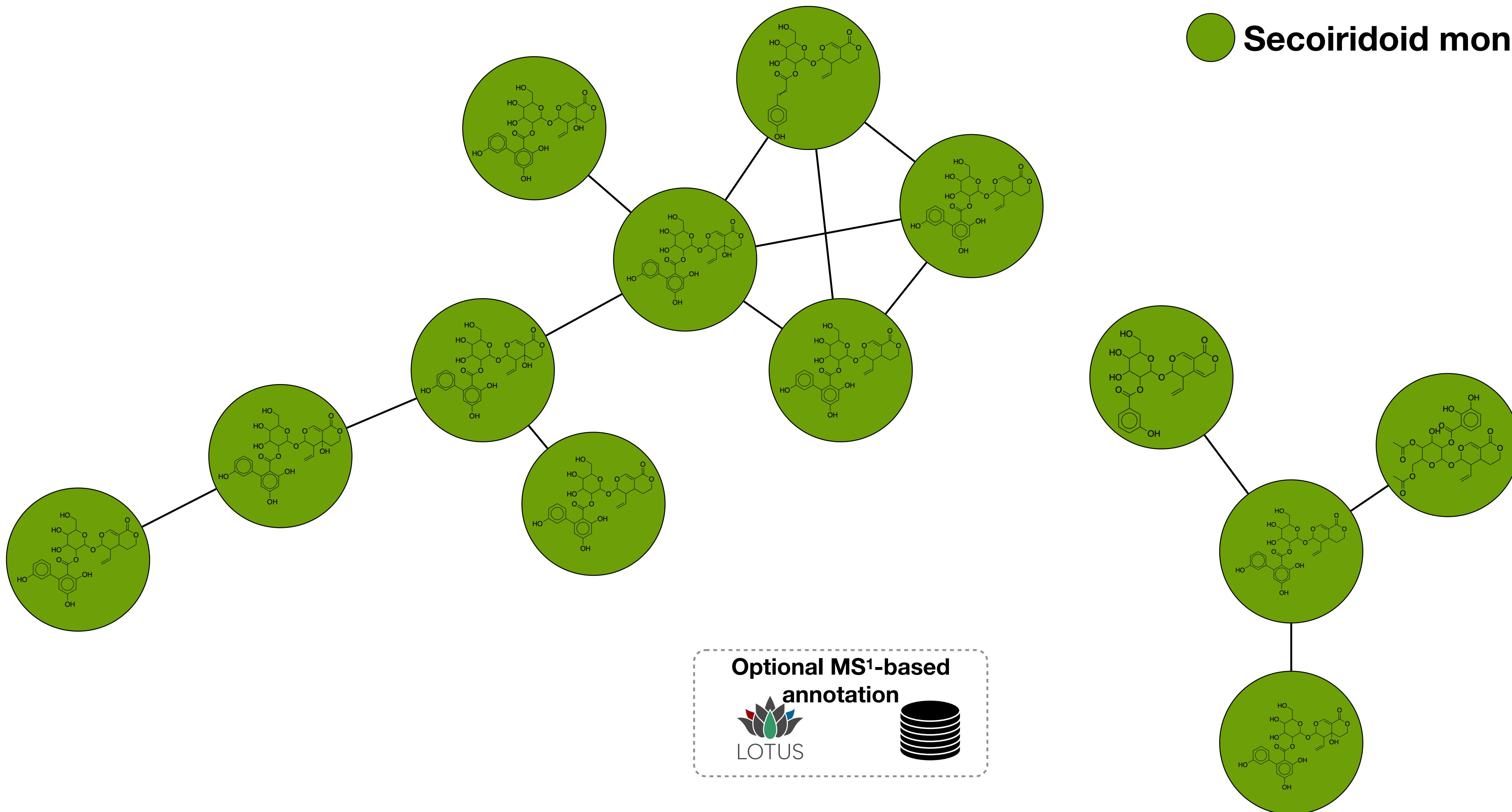
LOD-based annotation



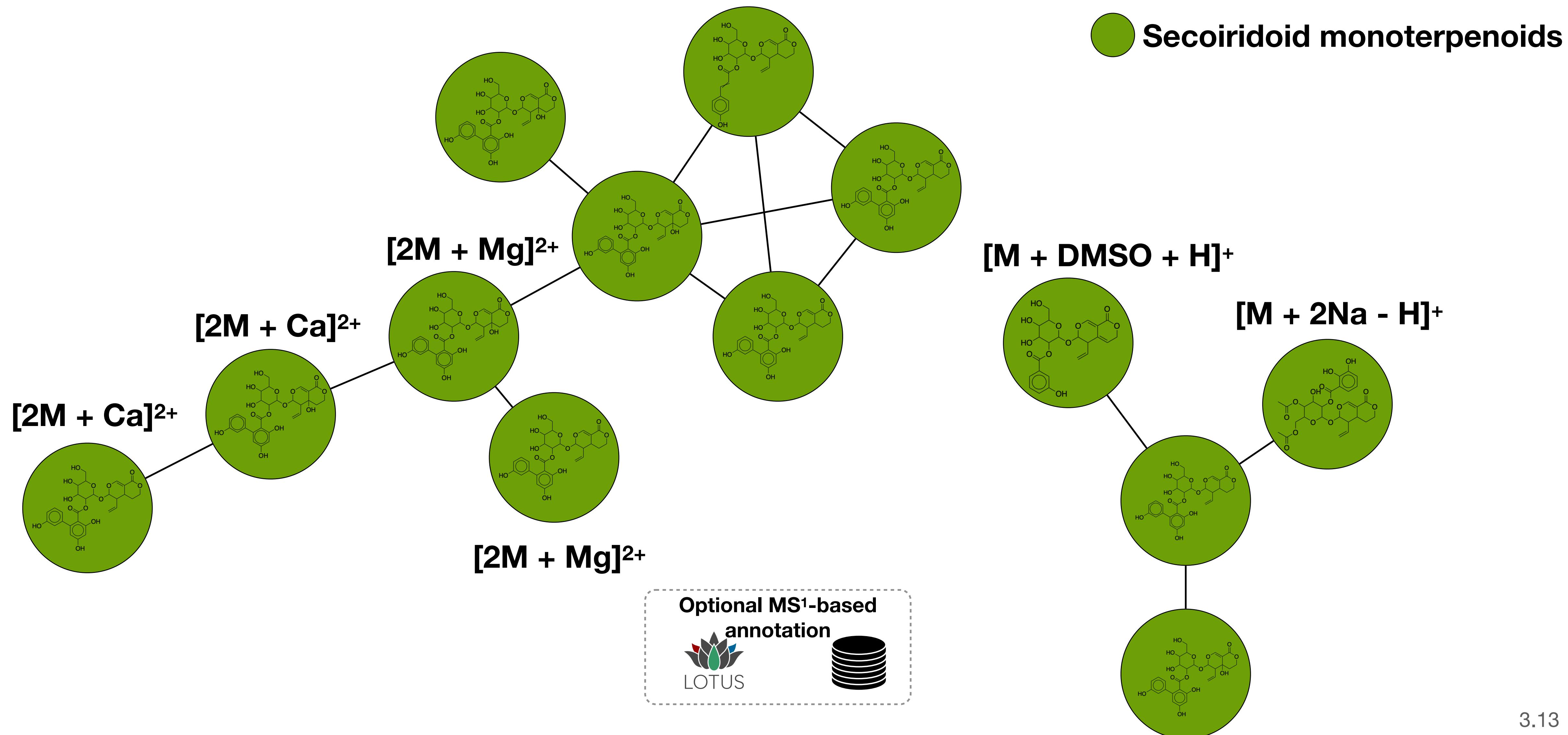
LOD-based annotation



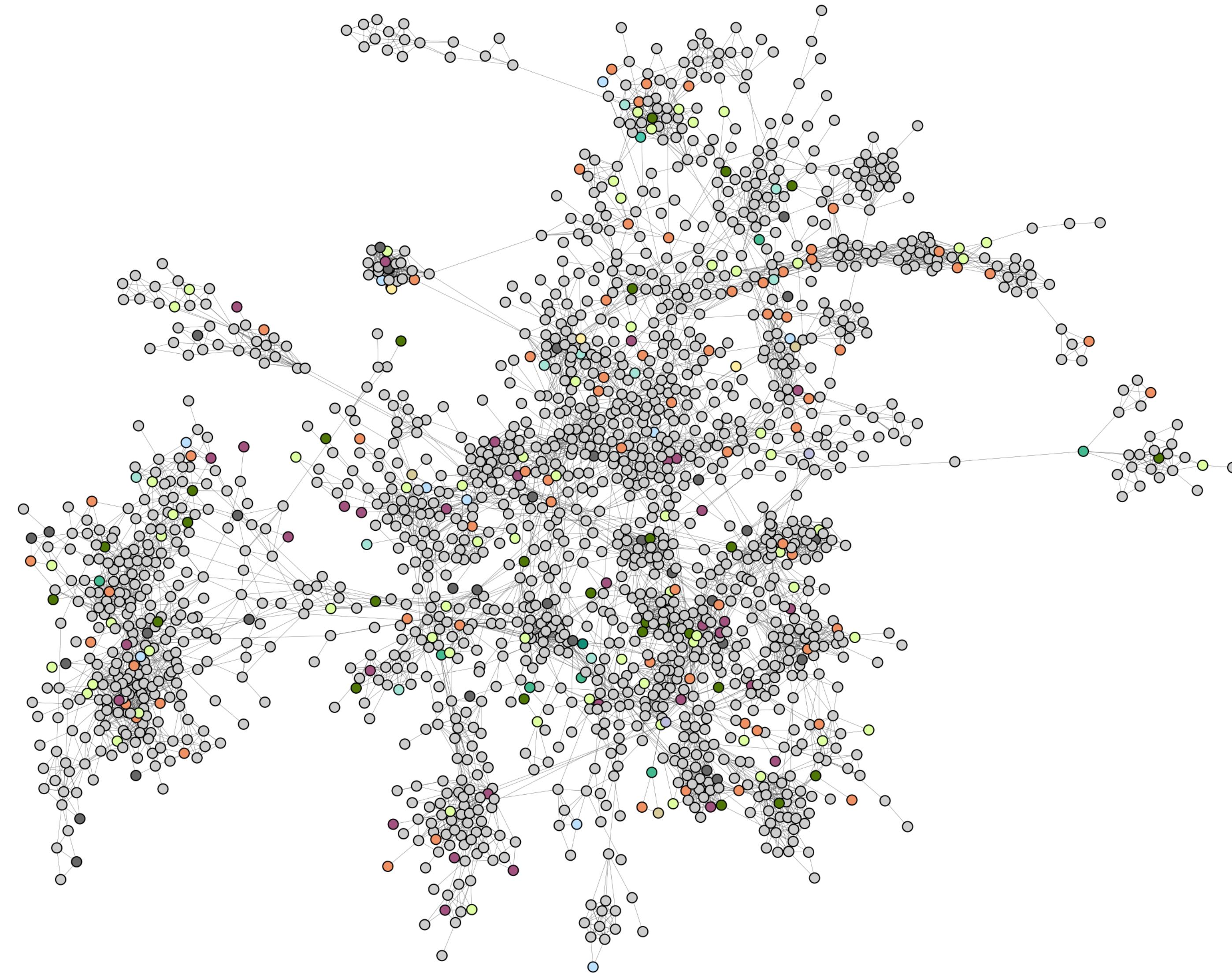
LOD-based annotation



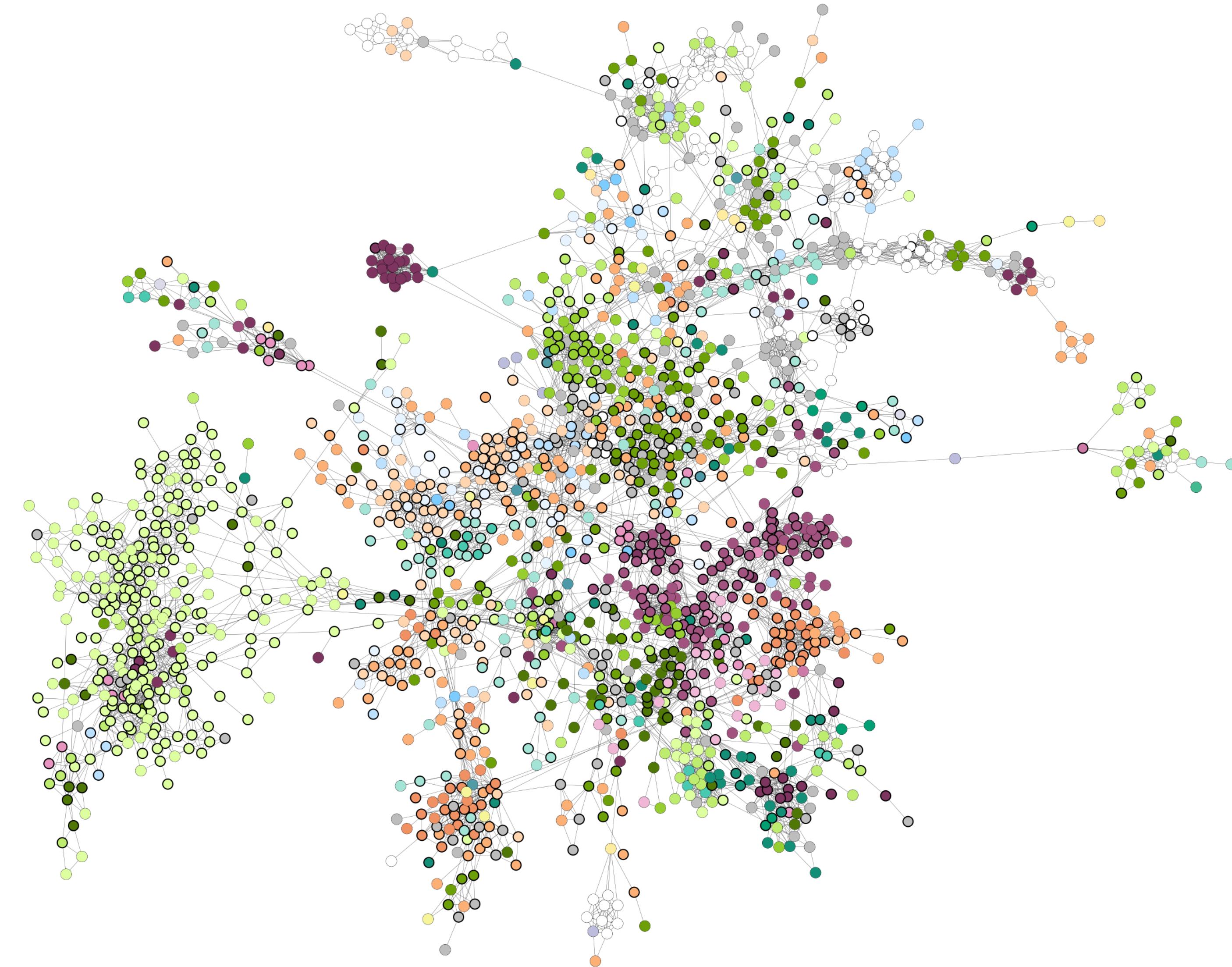
LOD-based annotation



Previous annotation



LOD-improved annotation



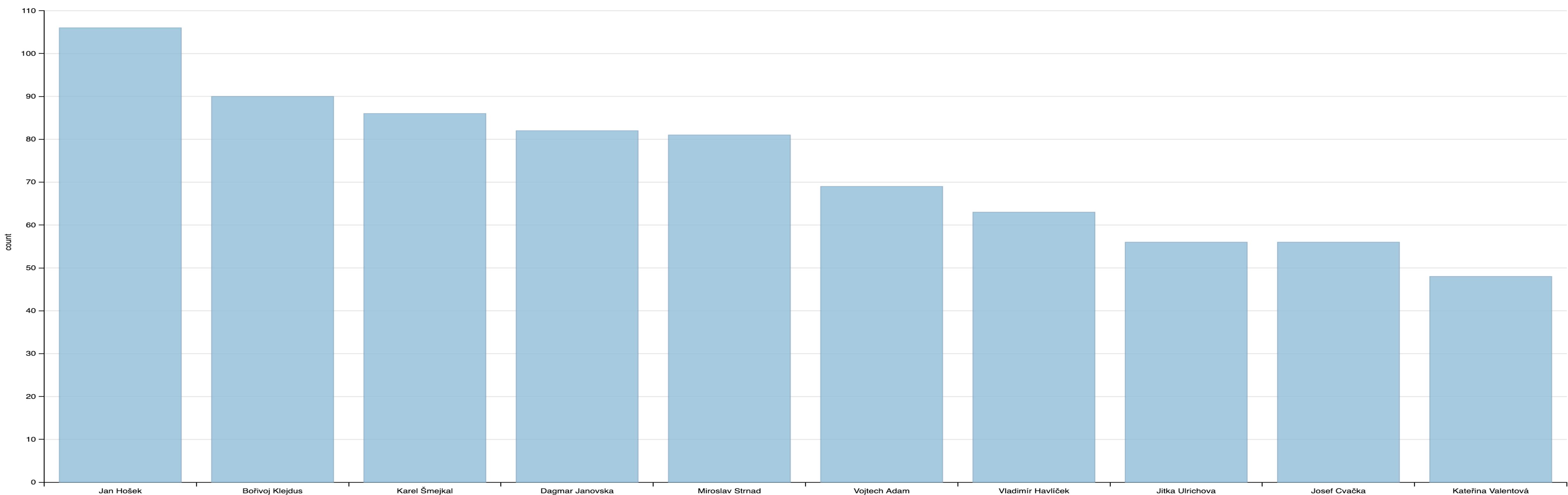
LOD has no limits

*« Which researchers employed by research institutions in the country of the city after which Pilsner (Q152281) is named published articles referencing compounds found in taxa?
(Top 10 compounds count) »*

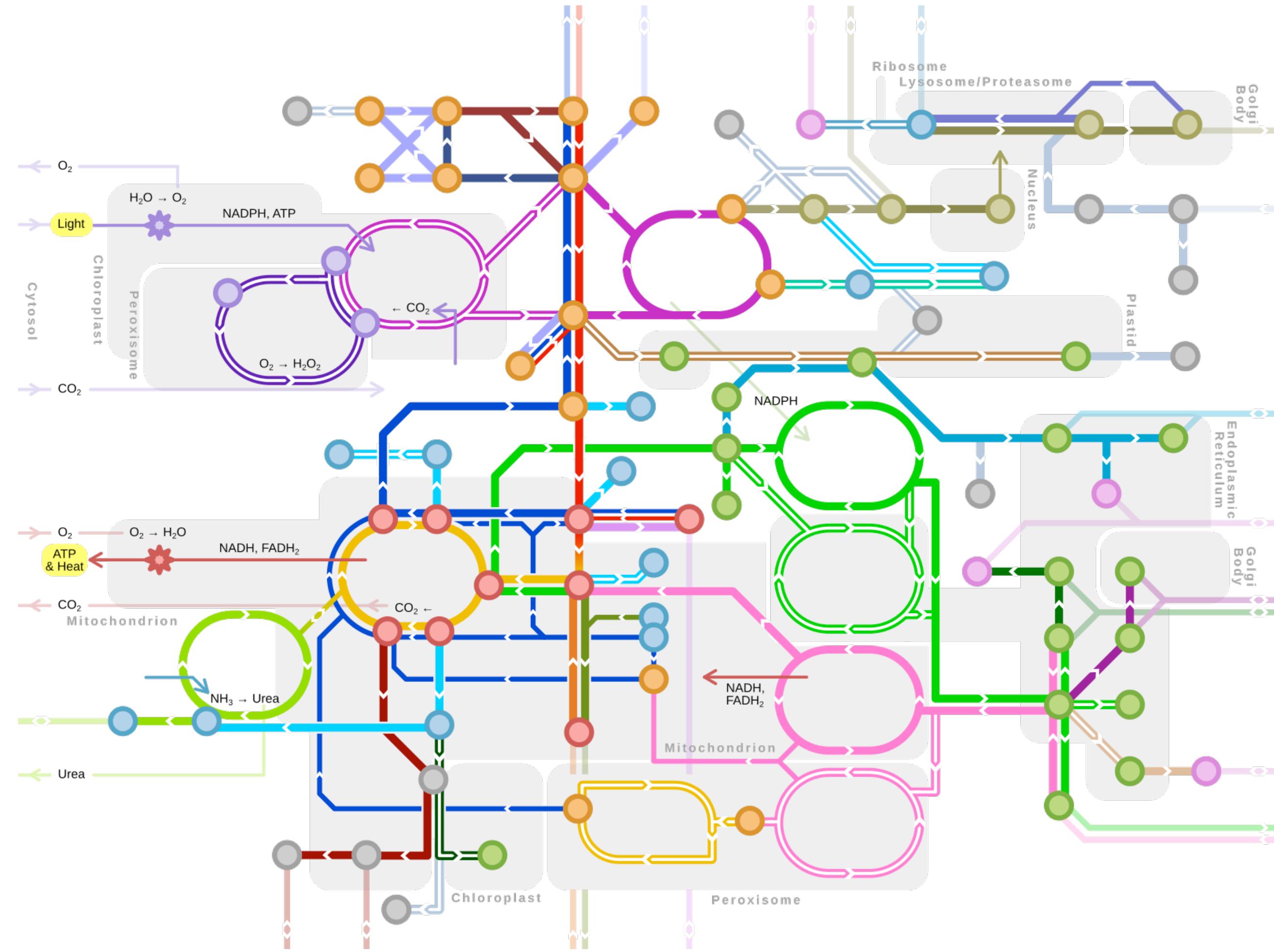


LOD has no limits

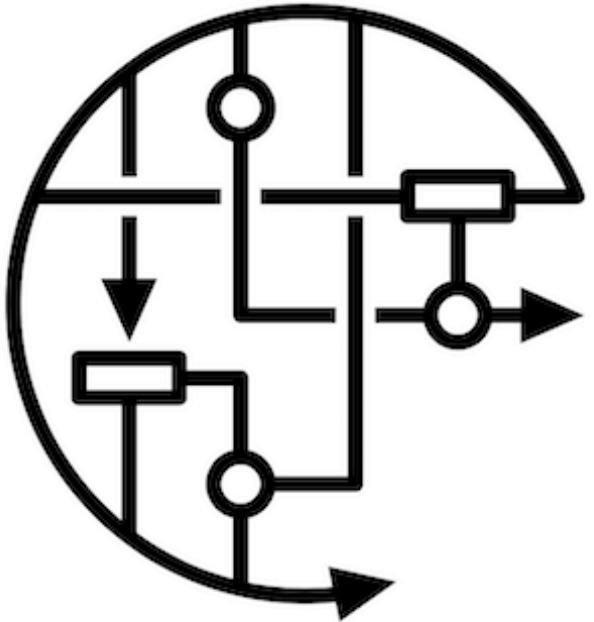
« Which researchers employed by research institutions in the country of the city after which Pilsner (Q152281) is named published articles referencing compounds found in taxa?
(Top 10 compounds count) »



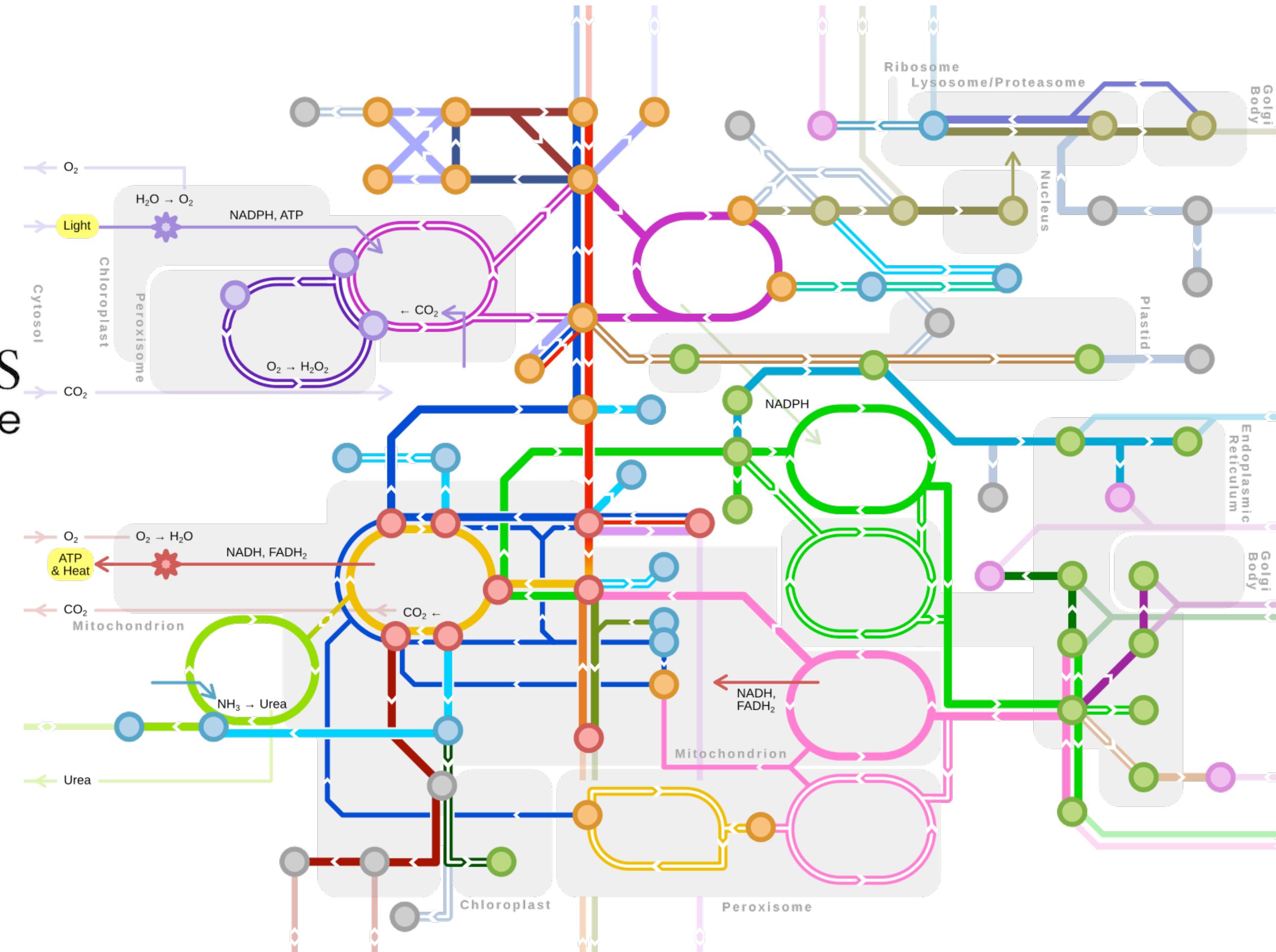
LOD has no limits



LOD has no limits



WIKIPATHWAYS
Pathways for the People



« Hey Wiki, are the metabolites I annotated known to interact with the cholesterol biosynthesis pathway? »

LOD has no limits

« Hey Wiki, I saw this review¹ yesterday, could you do the same?»



Contents lists available at [ScienceDirect](#)

Microbiological Research

journal homepage: www.elsevier.com/locate/micres



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose ^{a,*},¹ Anjisha Maharshi ^{a,2}, Bhavanath Jha ^{a,b,**}

^a Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

^b Academy of Scientific and Innovative Research (AcSIR), CSIR, India

¹ <https://doi.org/10.1016/j.micres.2021.126708>

LOD has no limits

« Hey Wiki, I saw this review¹ yesterday, could you do the same?»



Contents lists available at [ScienceDirect](#)

Microbiological Research

journal homepage: www.elsevier.com/locate/micres



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose ^{a,*},¹ Anjisha Maharshi ^{a,2}, Bhavanath Jha ^{a,b,**}

^a Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

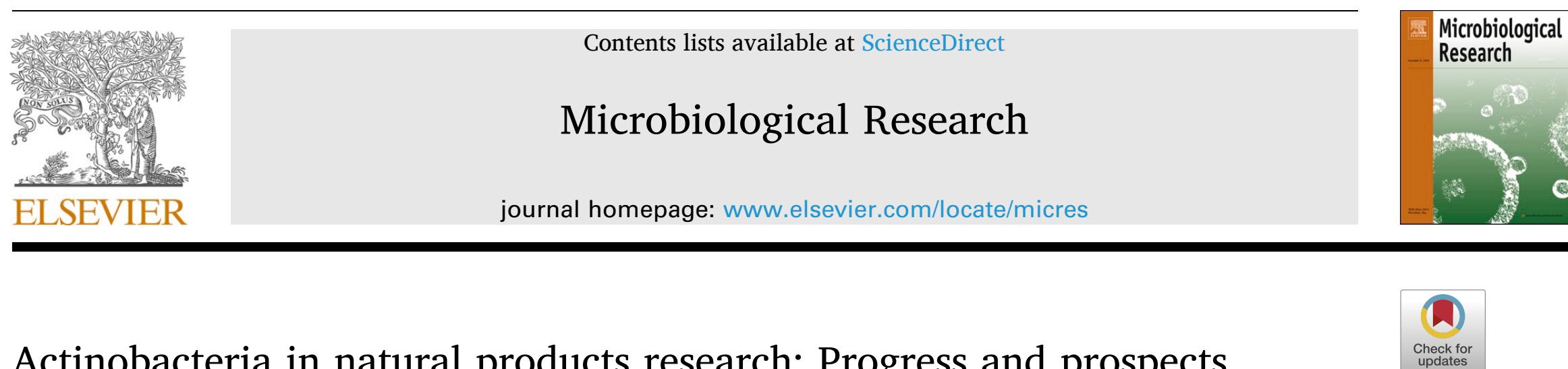
^b Academy of Scientific and Innovative Research (AcSIR), CSIR, India

Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?

¹ <https://doi.org/10.1016/j.micres.2021.126708>

LOD has no limits

« Hey Wiki, I saw this review¹ yesterday, could you do the same? »



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose ^{a,*}, Anjisha Maharshi ^{a,2}, Bhavanath Jha ^{a,b,**}

^a Marine Biotechnology and Ecology Division, CSIR- Central Salt and Marine Chemicals Research Institute, G. B. Marg, Bhavnagar, Gujarat, 364002, India

^b Academy of Scientific and Innovative Research (AcSIR), CSIR, India



Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?

508 results in 4.8 seconds

Table 1
Novel bioactive compounds isolated from Actinobacteria during last five years (2014-2019).

Organism	Compound	Bioactivity [#]	Reference
<i>Arthrobacter</i> sp. PGVB1	Arthroamide	f (anti-quorum sensing)	J Nat Prod. 2015;78:2827–2831
<i>Actinoallomorus</i> sp. ID145113 and 145,206	Paramagnetoquinones A-C (1-3)	(1/2) a; (3) a	J Nat Prod. 2017;80:819–827
<i>Actinoalloteichus hymeniacidonis</i> 179DD-027	Dokdolipid B	c	Mar Drugs 2019;17:237
<i>Actinomadura atramentaria</i> NBRC 14,695**	Cinnamycin B	a	J Ind Microbiol Biotechnol. 2016;43:1159–65
<i>Actinomadura</i> sp. BCC 35,430*	Actinomadurone	b	Tetrahedron Lett. 2017;58:3223–3225
<i>Actinomadura</i> sp. K13-0306	Sagamilactam	c, f (antiparasitic)	J Antibiot. 2016;69:818–824
<i>Actinomadura</i> sp. K4S16	Nonthmicin (1) & ecteinamycin (2)	(1,2) c	Org Lett. 2017;19:1406–1409
<i>Actinomadura</i> sp. KC191*	Actinomadurol	a	J Nat Prod. 2016;79:1886–1890
<i>Actinosynnema pretiosum</i> HGF052::asm18**	Actinosynneptide A (1) & B (2)	(1,2) c	Appl Microbiol Biotechnol. 2016;101:2273–2279
<i>Amycolatopsis</i> sp. IRD-009	Pradimicin-IRD	a, c	Natural Product Res. 2019;33:1713–1720

Taxon name	Structure	Activity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic immunosuppressive drug	10.1007/S11306-016-025-6
<i>Saccharopolyspora erythraea</i>		Antibiotic protein synthesis inhibitors gastrointestinal agent	10.1021/ACSSYNBIO.8B00372
<i>Streptomyces venezuelae</i>		Antibiotic protein synthesis inhibitors	10.1016/J.JBIOTEC.2014.01.028

¹ <https://doi.org/10.1016/j.micres.2021.126708>

LOD has (almost) no limits

« Hey Wiki, I saw this review¹ yesterday, could you do the same? »



Actinobacteria in natural products research: Progress and prospects

Polpass Arul Jose ^{a,*},¹ Anjisha Maharshi ^{a,2}, Bhushan Patel ^{a,2,*,**}

^a Marine Biotechnology and Ecology Division, CSIR-Central Salt and Marine Chemicals Research Institute, Regional Centre, Bhavnagar Marg, Bhavnagar, Gujarat, 364002, India

^b Academy of Scientific and Innovative Research (AcSIR), CSIR, India

Table 1
Novel bioactive compounds isolated from Actinobacteria during last five years (2014–2019).

Organism	Compound	Bioactivity [#]	Reference
<i>Arthrobacter</i> sp. PGVB1	Arthroamide	f (anti-quorum sensing)	J Nat Prod. 2015;78:2827–2831
<i>Actinoallomorus</i> sp. ID145113 and 145,206	Paramagnetoquinones A–C (1–3)	(1/2) a; (3) a	J Nat Prod. 2017;80:819–827
<i>Actinoalloteichus hymeniacidonis</i> 179DD-027	Dokdolipid B	c	Mar Drugs 2019;17:237
<i>Actinomadura atramentaria</i> NBRC 14,695 ^{**}	Cinnamycin B	a	J Ind Microbiol Biotechnol. 2016;43:1159–65
<i>Actinomadura</i> sp. BCC 35,430 [*]	Actinomadurone	b	Tetrahedron Lett. 2017;58:3223–3225
<i>Actinomadura</i> sp. K13–0306	Sagamilactam	c, f (antiparasitic)	J Antibiot. 2016;69:818–824
<i>Actinomadura</i> sp. K4S16	Nonthmicin (1) & ecteinamycin (2)	(1,2) c	Org Lett. 2017;19:1406–1409
<i>Actinomadura</i> sp. KC191 [*]	Actinomadurol	a	J Nat Prod. 2016;79:1886–1890
<i>Actinosynnema pretiosum</i> HGF052::asm18 ^{**}	Actinosynneptide A (1) & B (2)	(1,2) c	Appl Microbiol Biotechnol. 2016;101:2273–2279
<i>Amycolatopsis</i> sp. IRD-009	Pradimicin-IRD	a, c	Natural Product Res. 2019;33:1713–1720

Which compounds with known bioactivities were isolated from Actinobacteria (Q26262282), between 2014 and 2019, with related organisms and references?

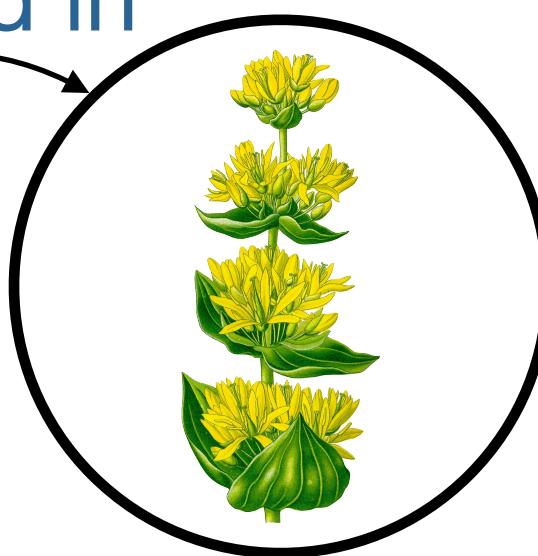
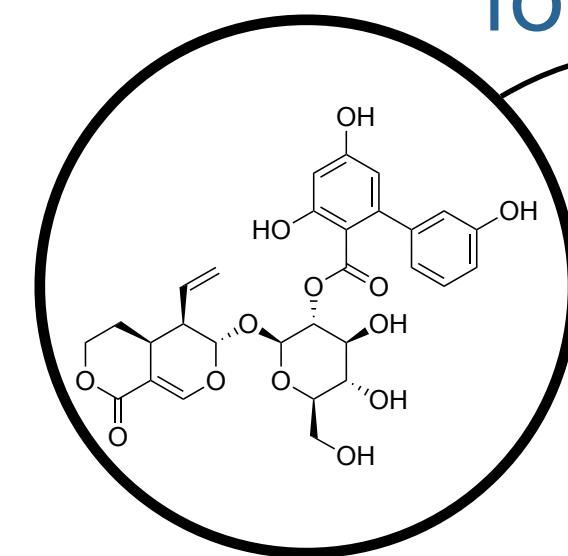
Taxon name	Structure	Bioactivity	Reference
<i>Streptomyces coelicolor</i>		Antibiotic immunosuppressive drug	10.1007/S11306-016-025-6
<i>Saccharopolyspora erythraea</i>		Antibiotic protein synthesis inhibitors gastrointestinal agent	10.1021/ACSSYNBIO.8B00372
<i>Streptomyces venezuelae</i>		Antibiotic protein synthesis inhibitors	10.1016/J.JBIOTEC.2014.01.028

¹ <https://doi.org/10.1016/j.micres.2021.126708>

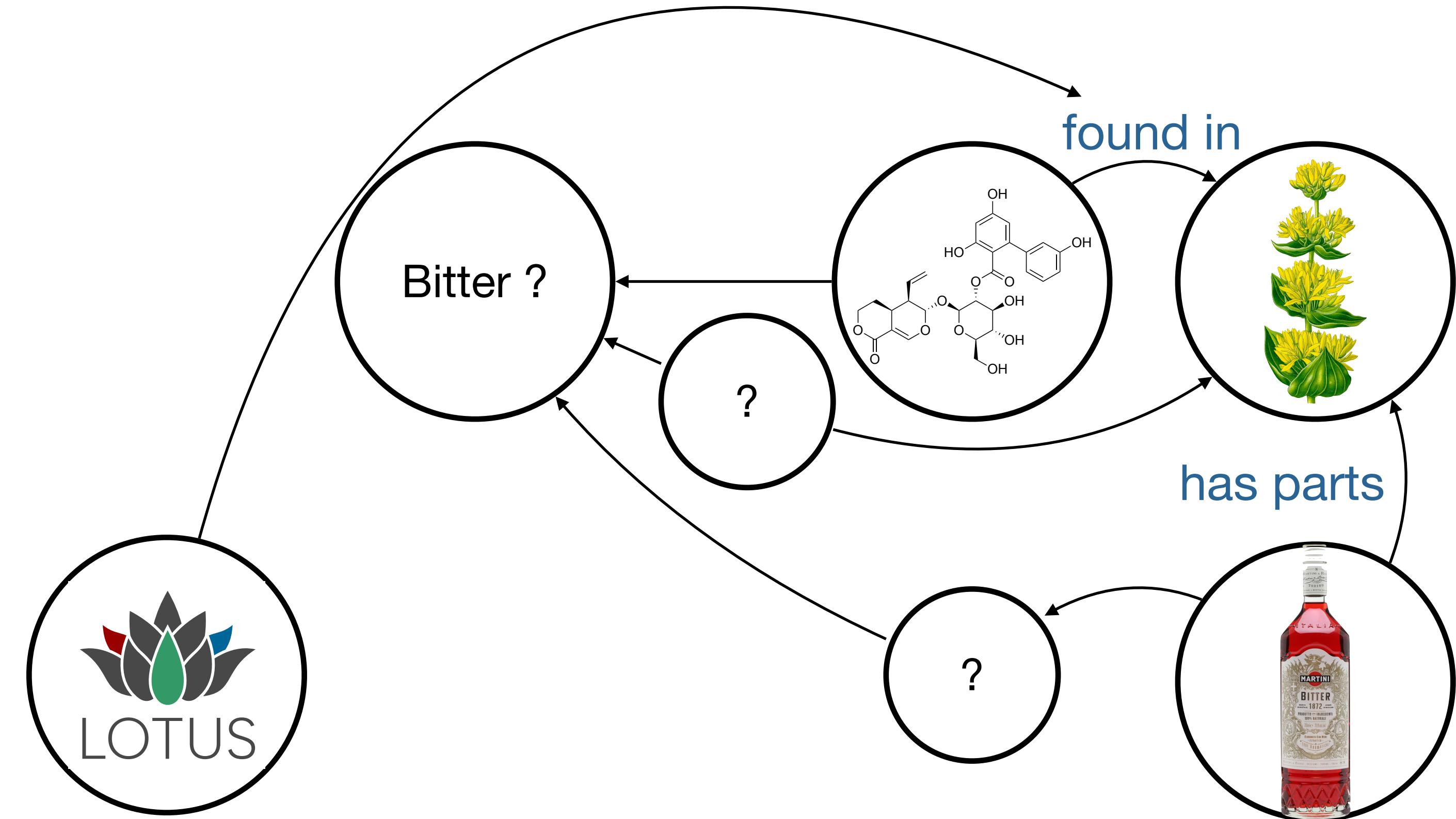
LOD has no limits



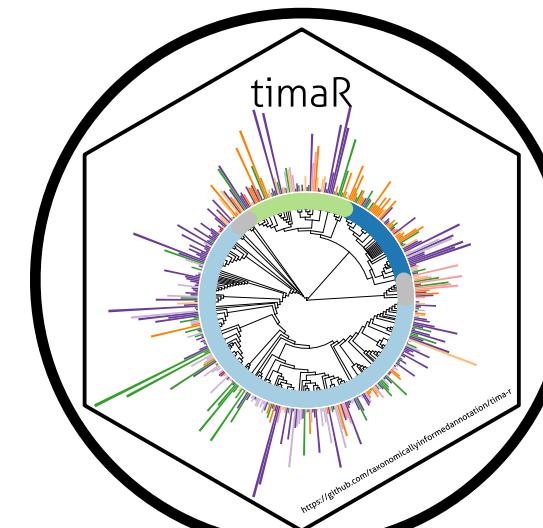
found in



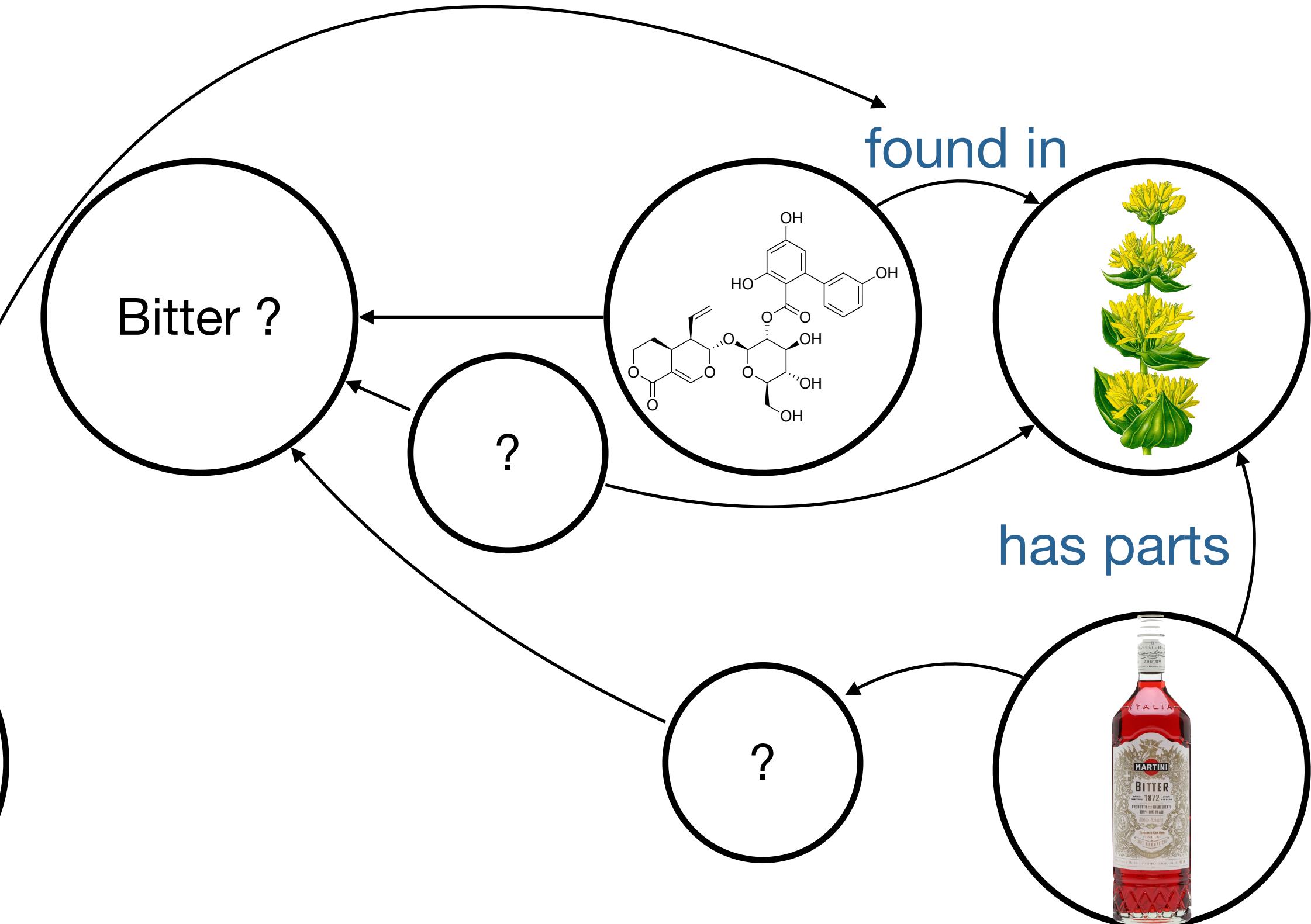
LOD has no limits



LOD has no limits



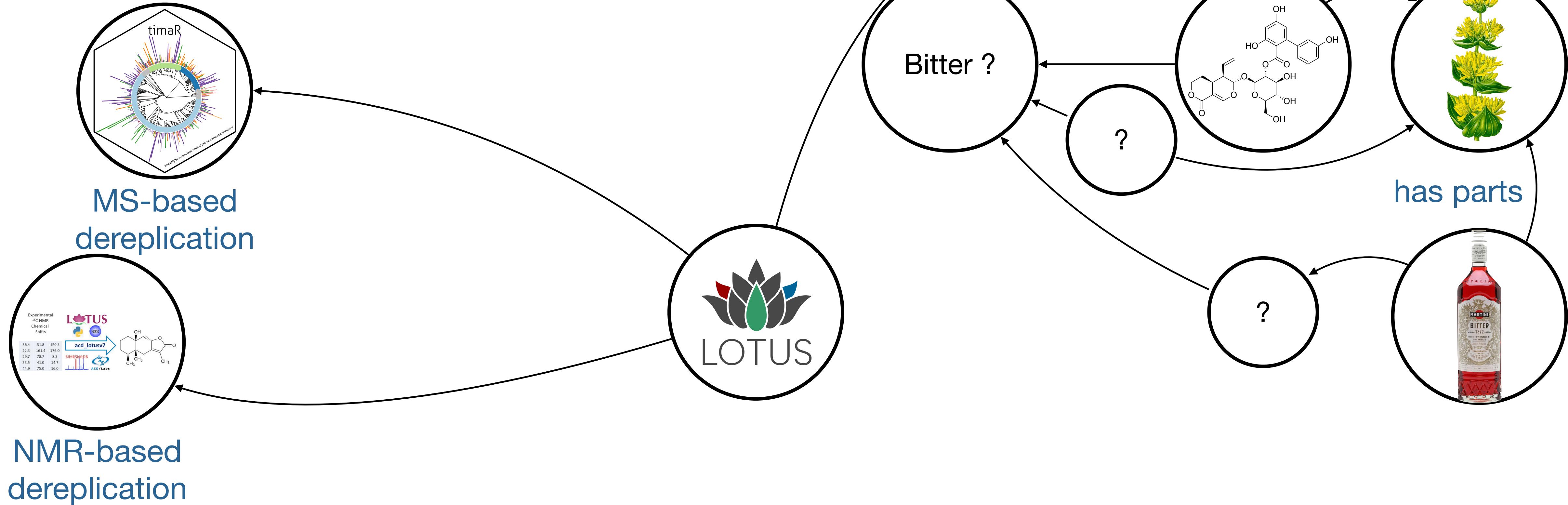
MS-based
dereplication



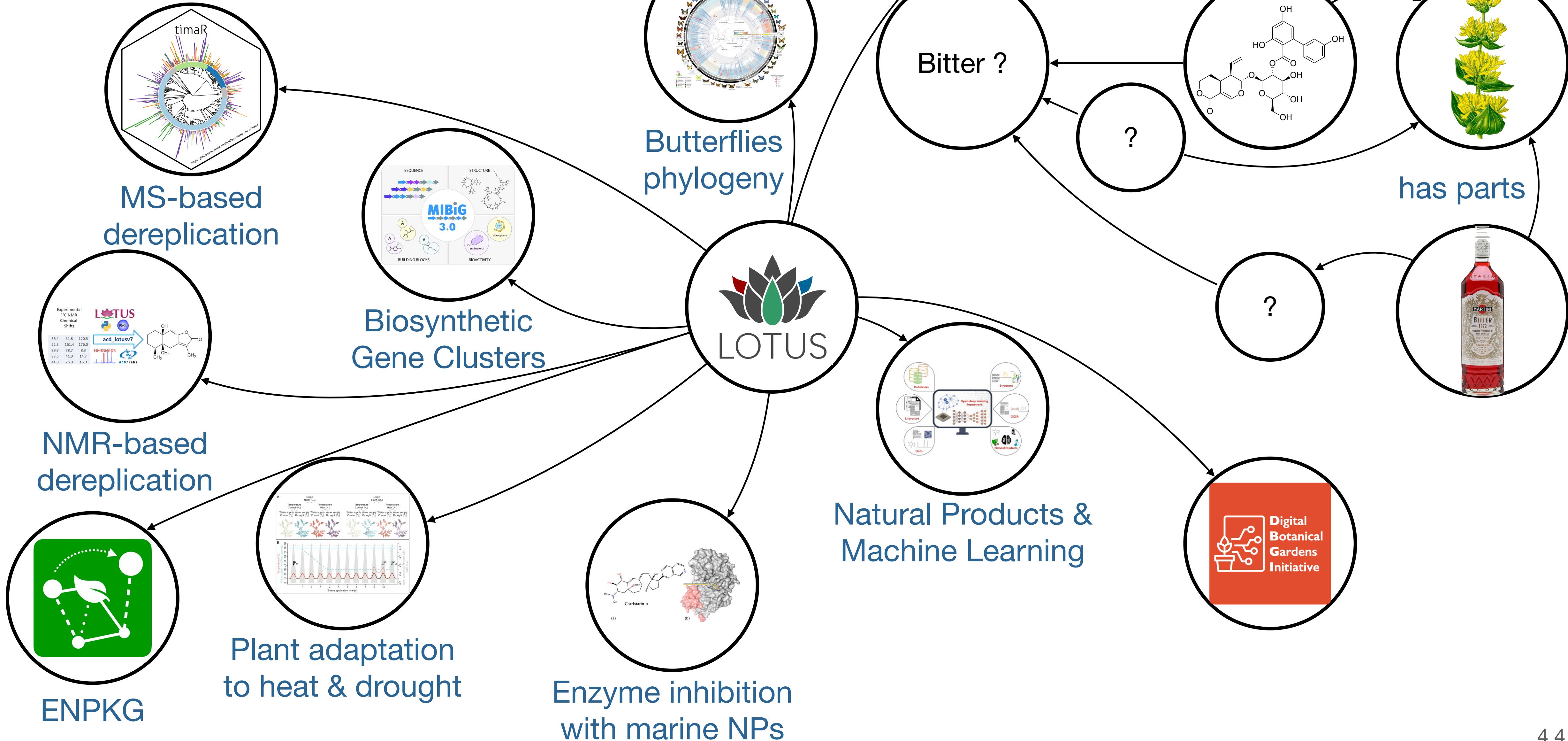
has parts



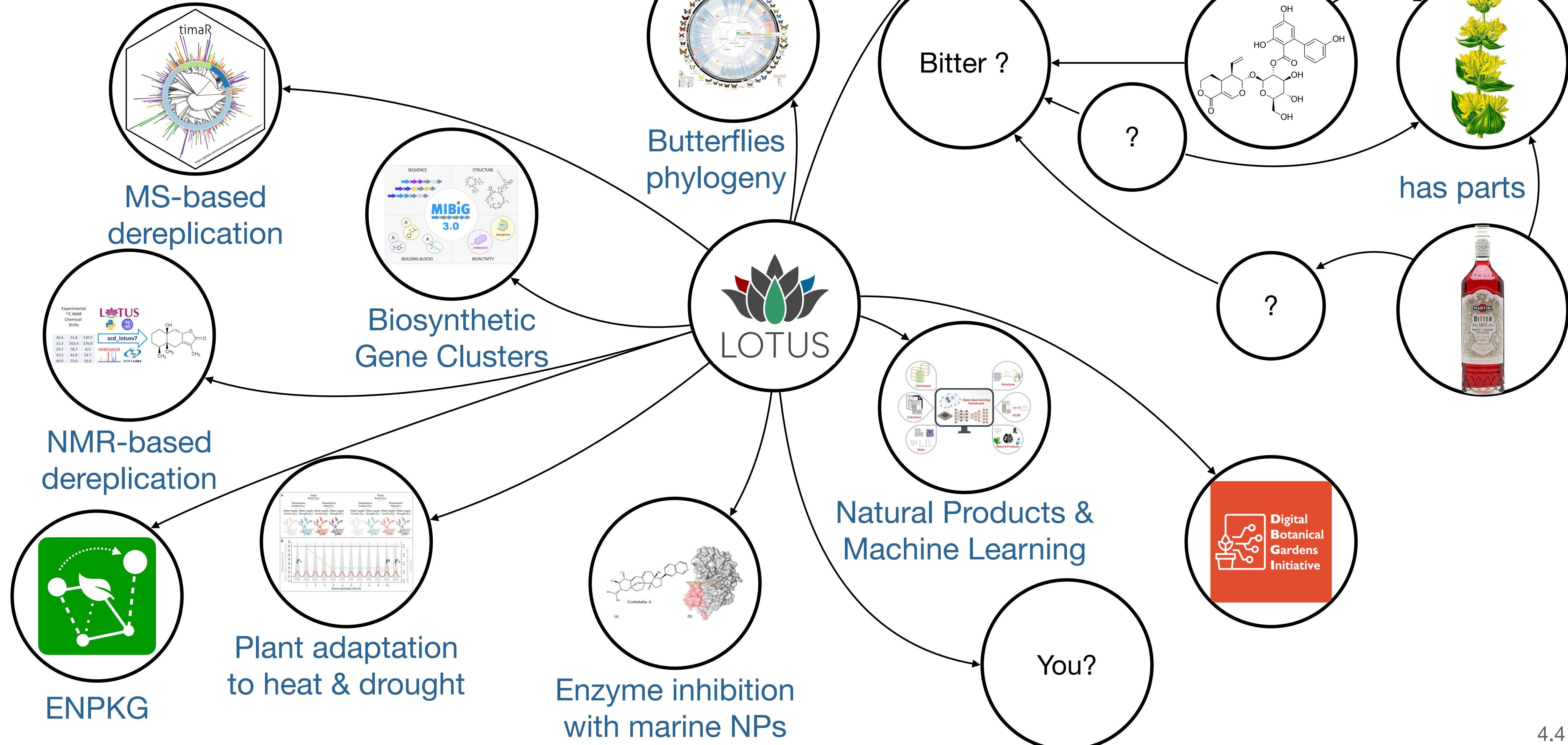
LOD has no limits



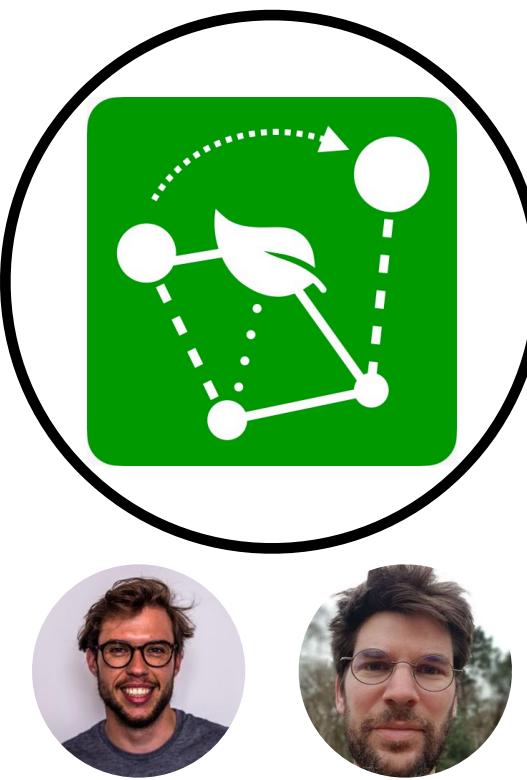
LOD has no limits



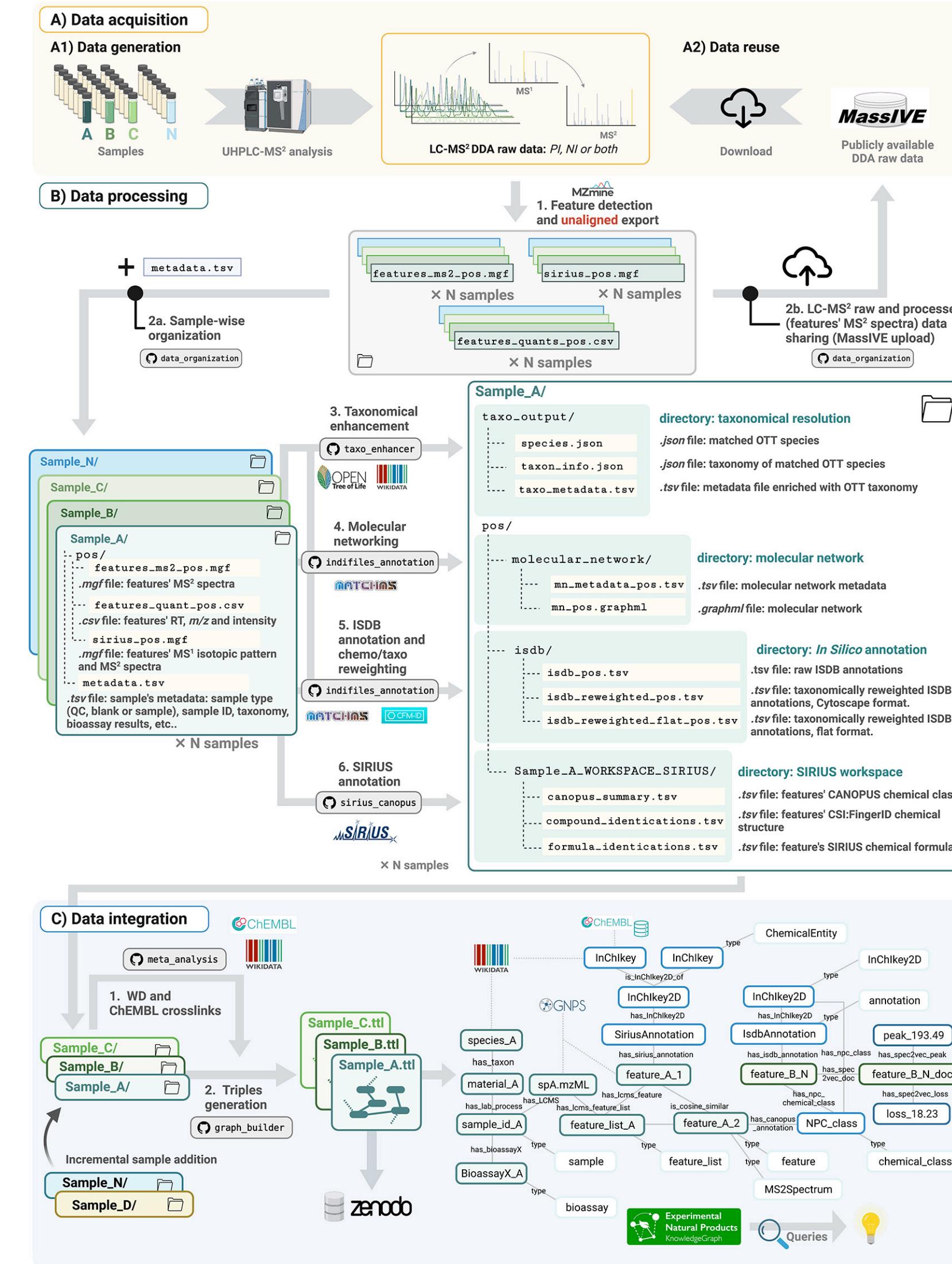
LOD has no limits



The Experimental Natural Products Knowledge Graph (ENPKG)



<https://enpkg.commonslab.org/>



For more: <https://doi.org/10.1021/acscentsci.3c00800>

Acknowledgements

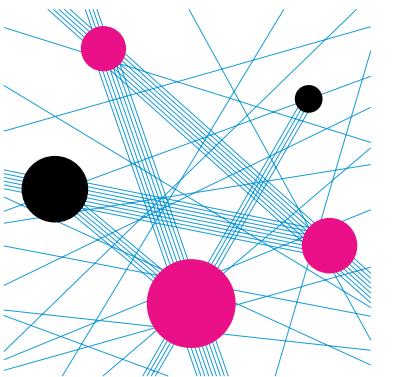


ETH zürich
D BIOL

Institute of Molecular Systems Biology



Prof. Nicola Zamboni



Contact



rutz@imsb.biol.ethz.ch