

UNIVERSITÉ
DE GENÈVE

FACULTY OF SCIENCE

Section of Pharmaceutical Sciences

The LOTUS Initiative for Knowledge Sharing in Natural Products Research

69th International Congress and Annual Meeting
of the Society for Medicinal Plant and
Natural Product Research (GA)

Online meeting, 2021-09-08

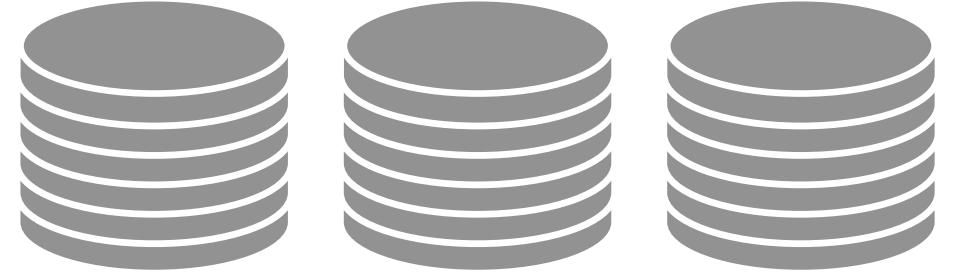
The team



The team



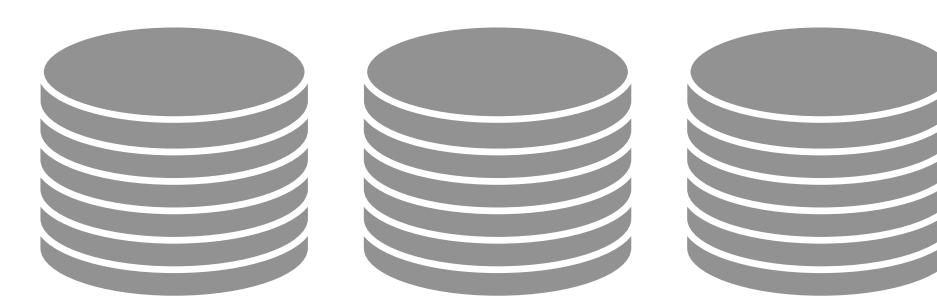
The initiative



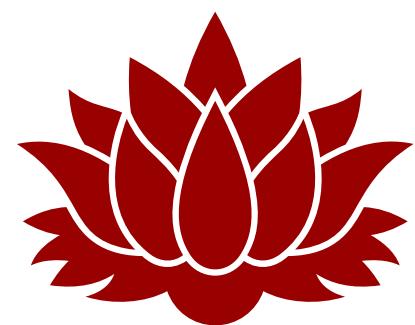
Available Natural
Product Data

Unstructured,
redundant and
noisy data

The initiative



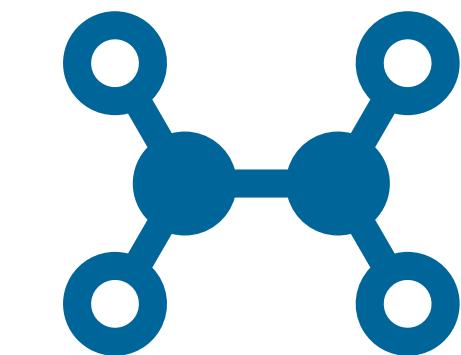
Available Natural
Product Data



Biological
Taxon



Reference

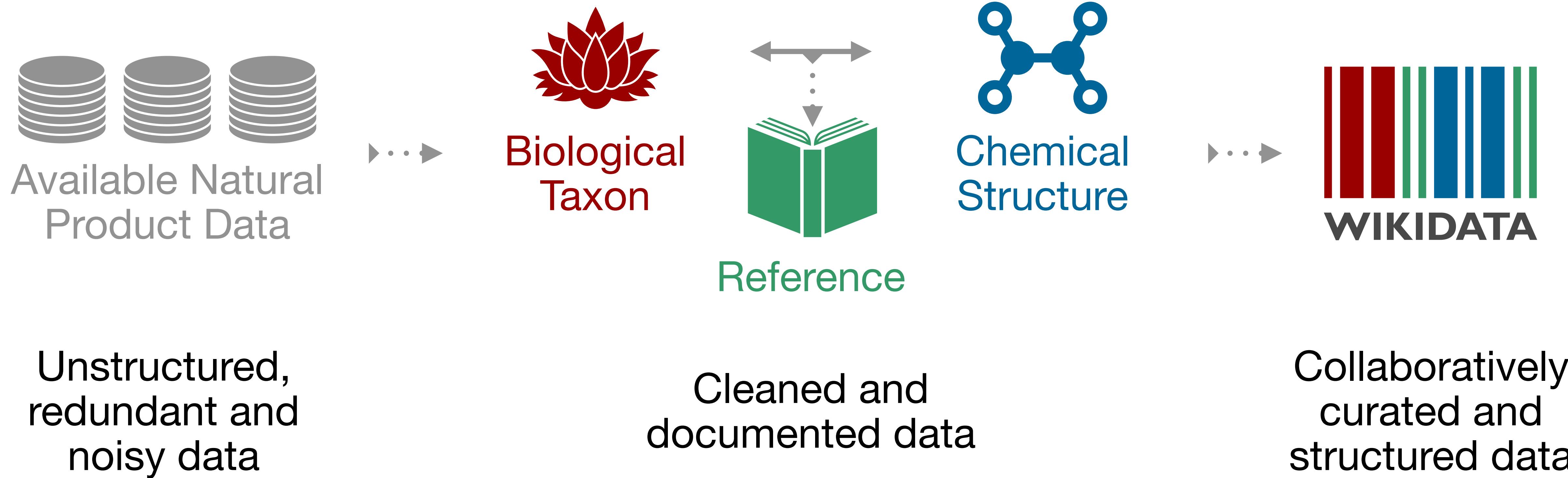


Chemical
Structure

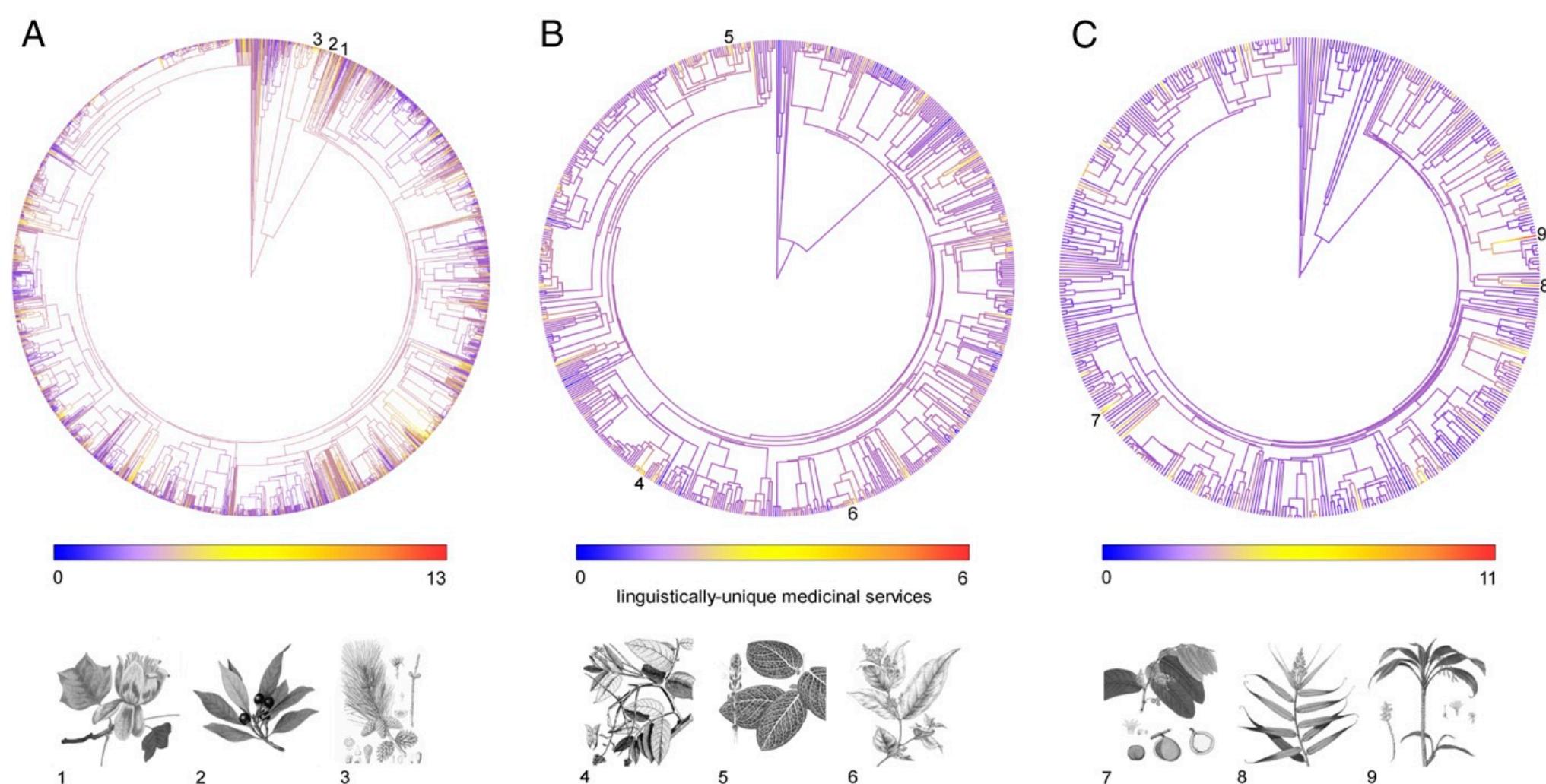
Unstructured,
redundant and
noisy data

Cleaned and
documented data

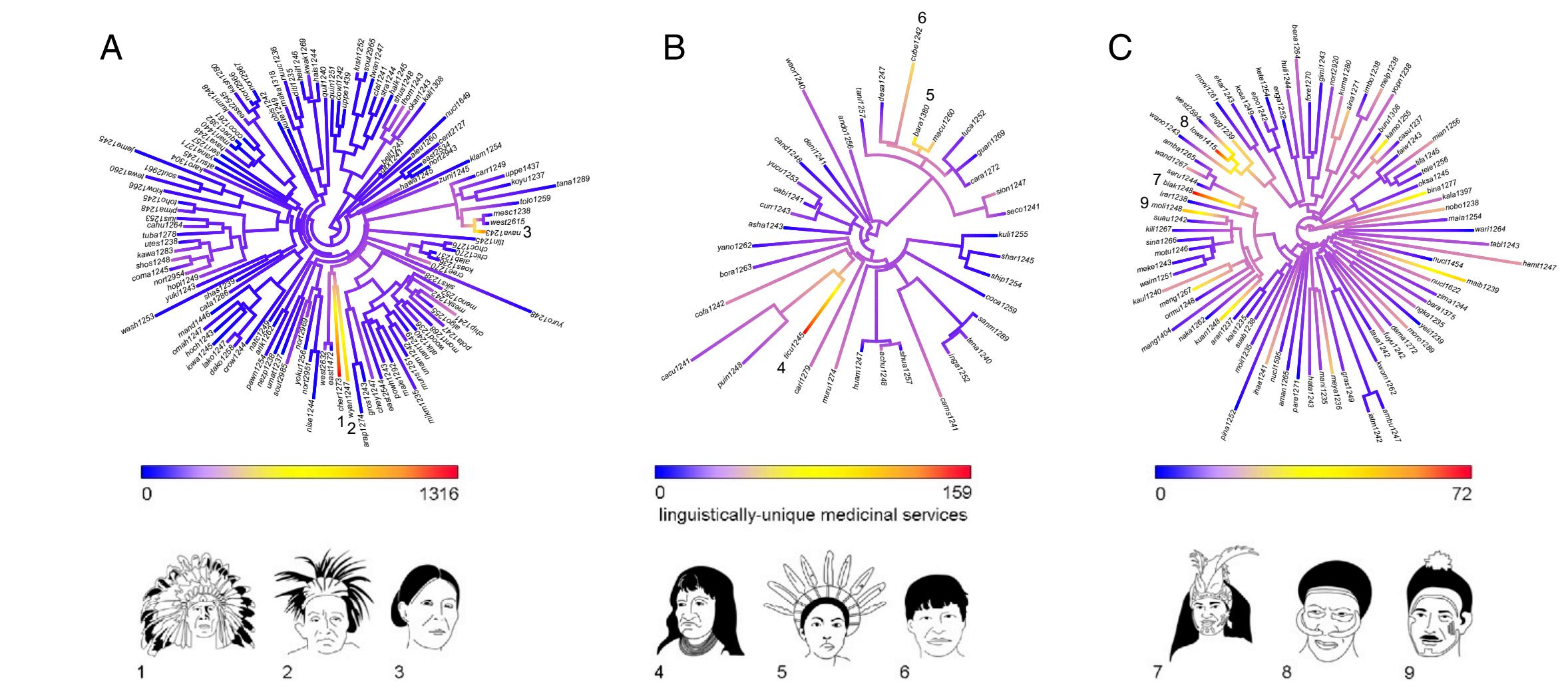
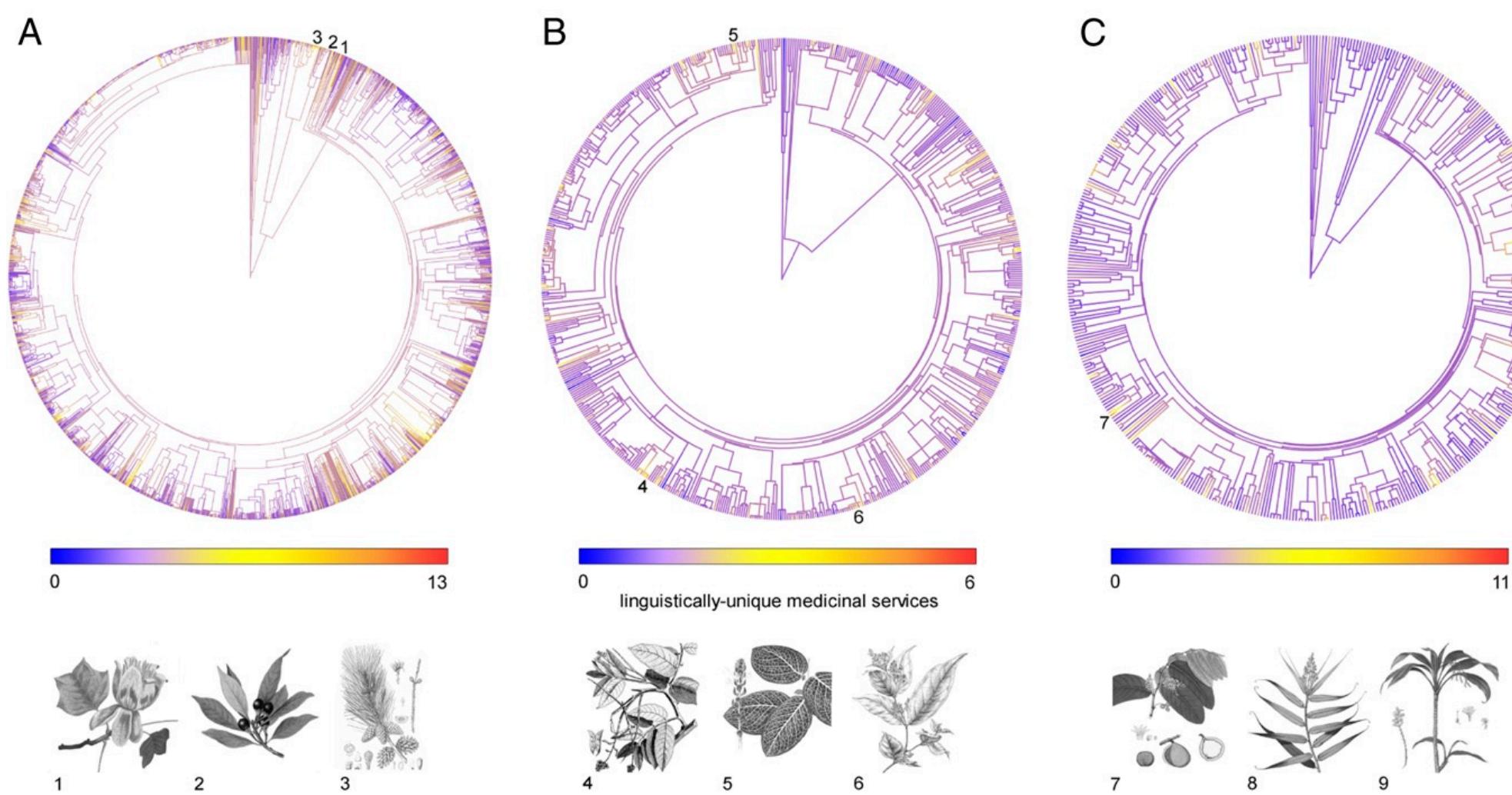
The initiative



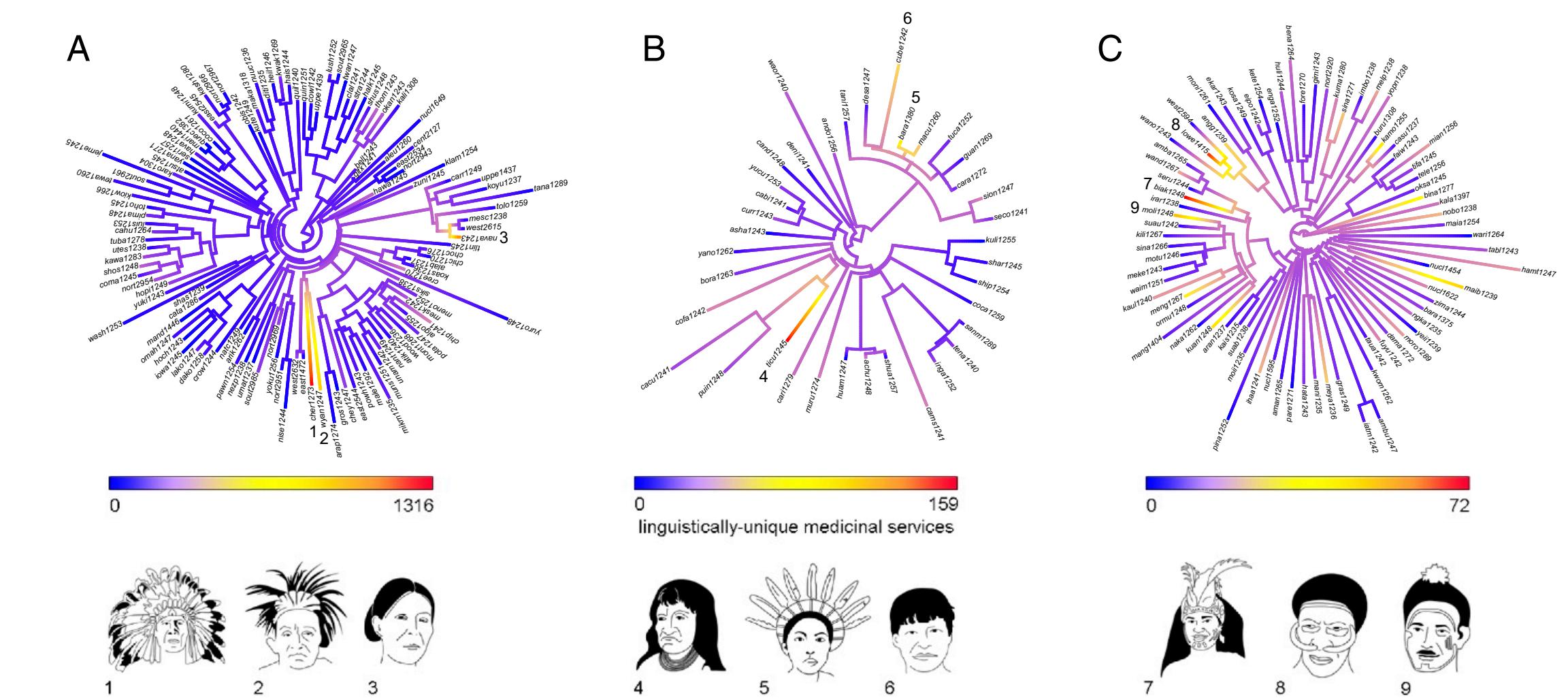
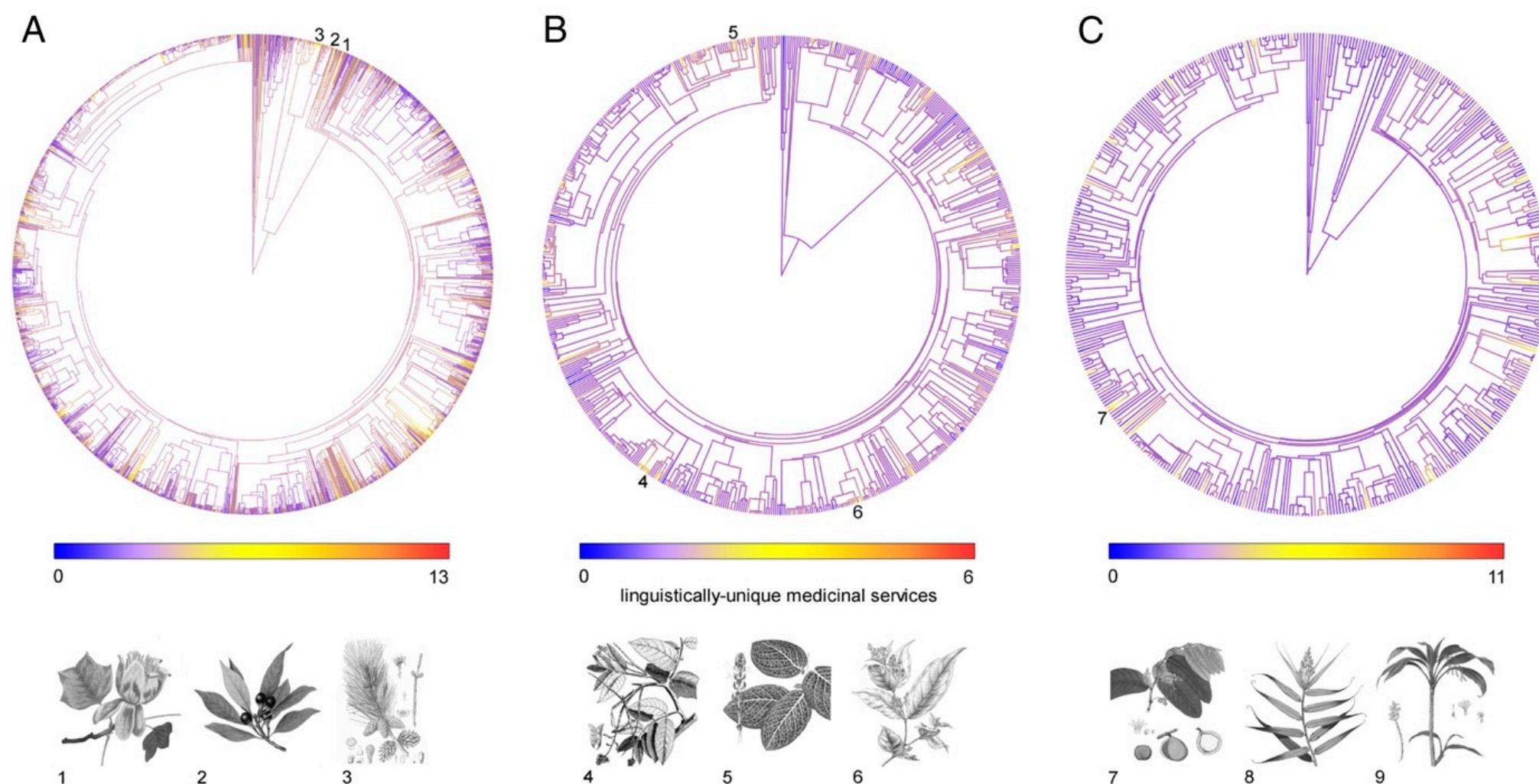
The initiative - *Why?*



The initiative - *Why?*

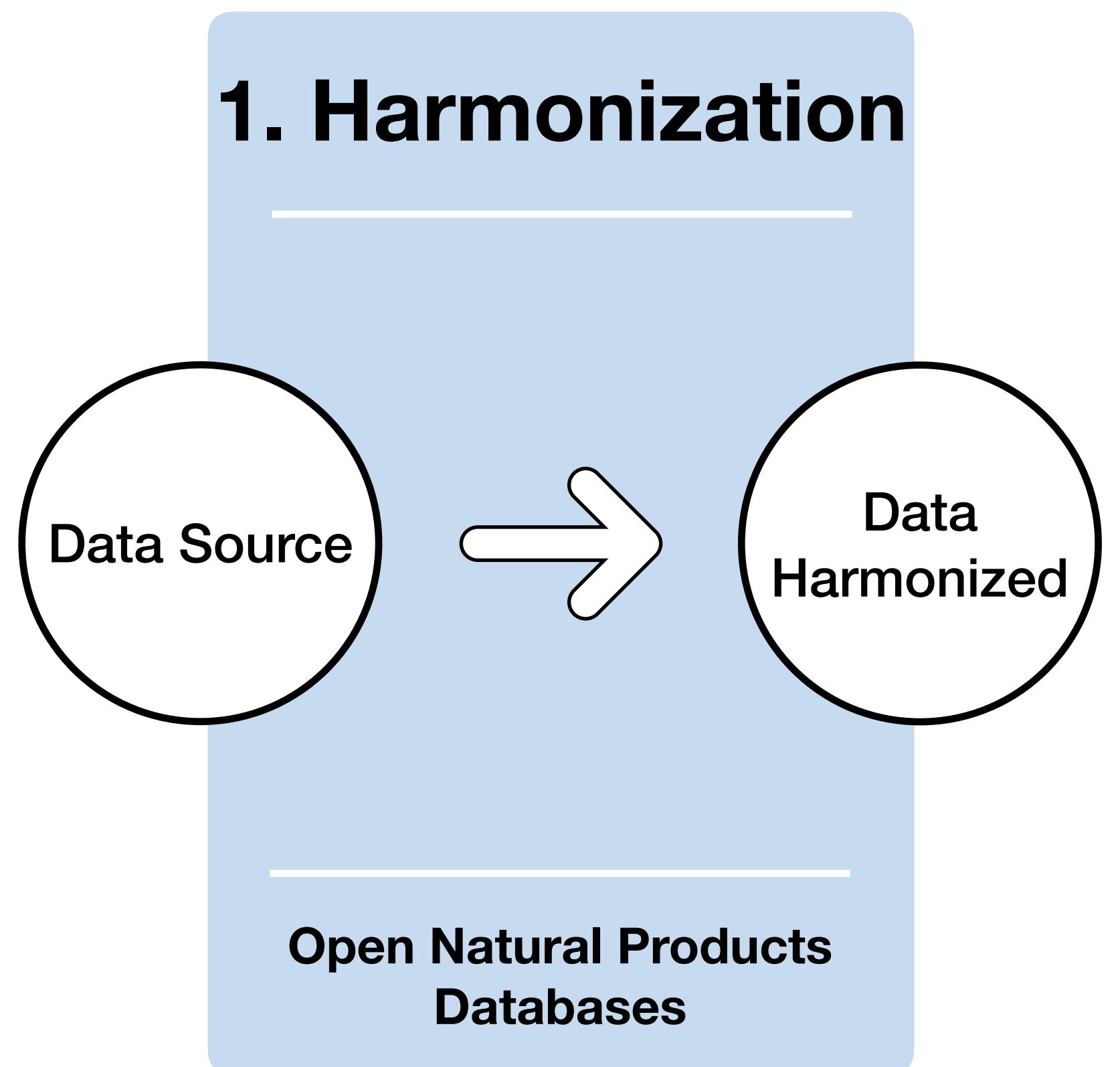


The initiative - *Why?*

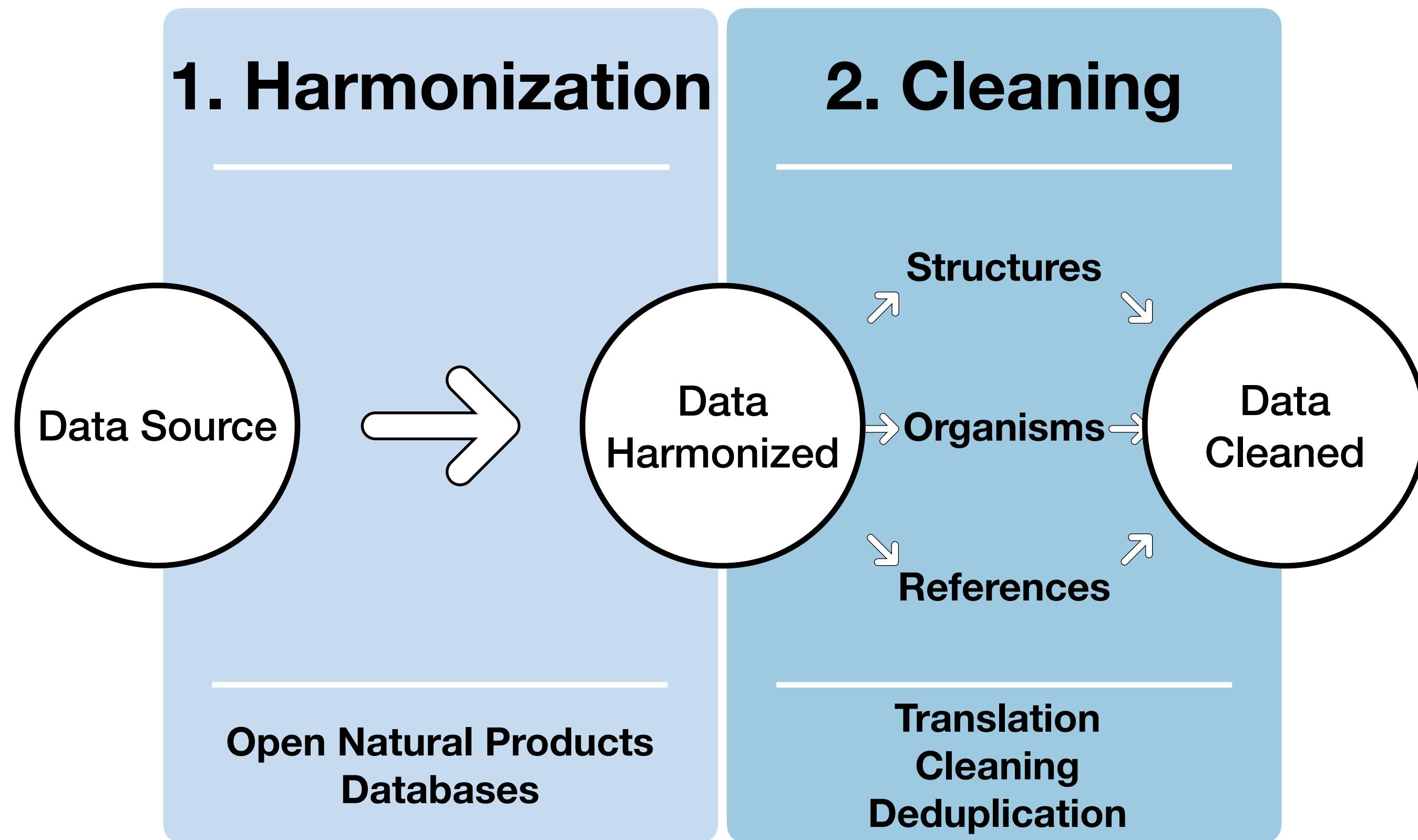


« Our finding of high uniqueness in indigenous knowledge and strong coupling with threatened languages suggests that language loss will be even more critical to the extinction of medicinal knowledge than biodiversity loss »

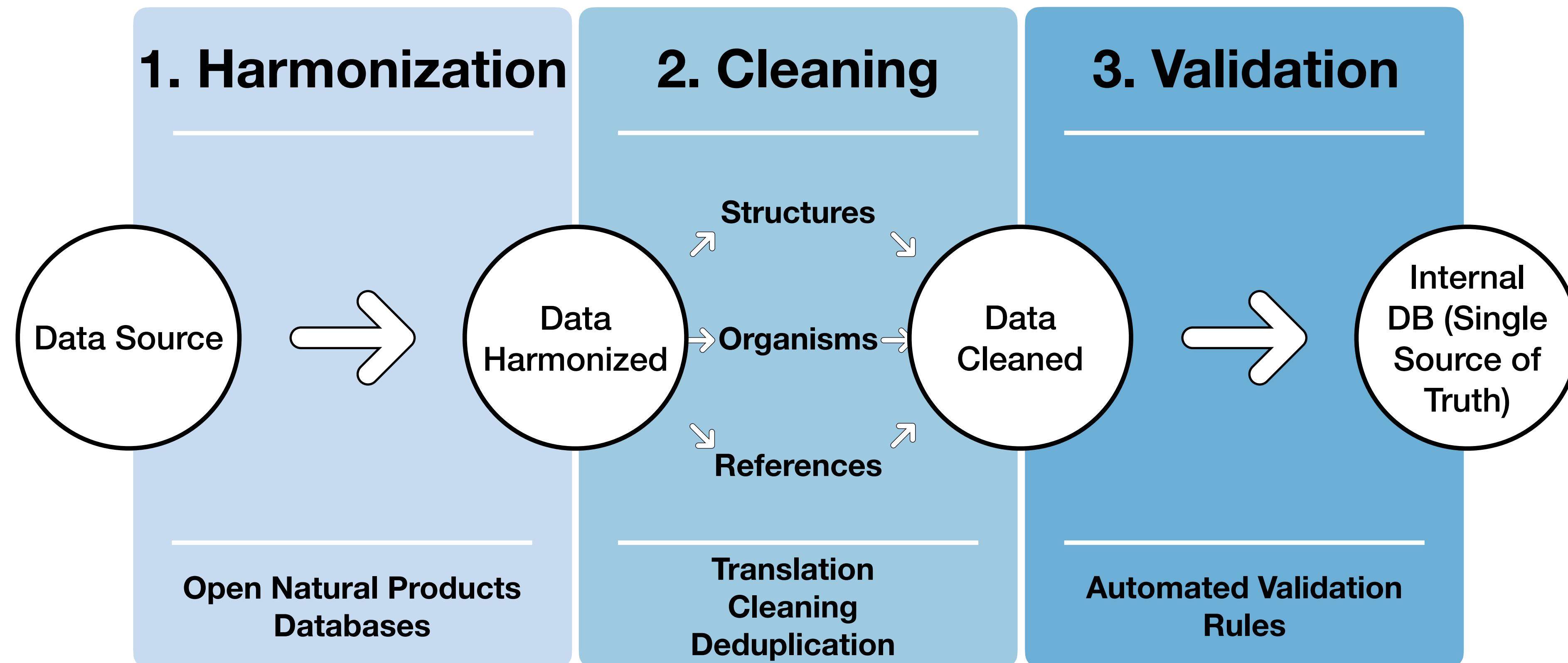
The initiative - *How?*



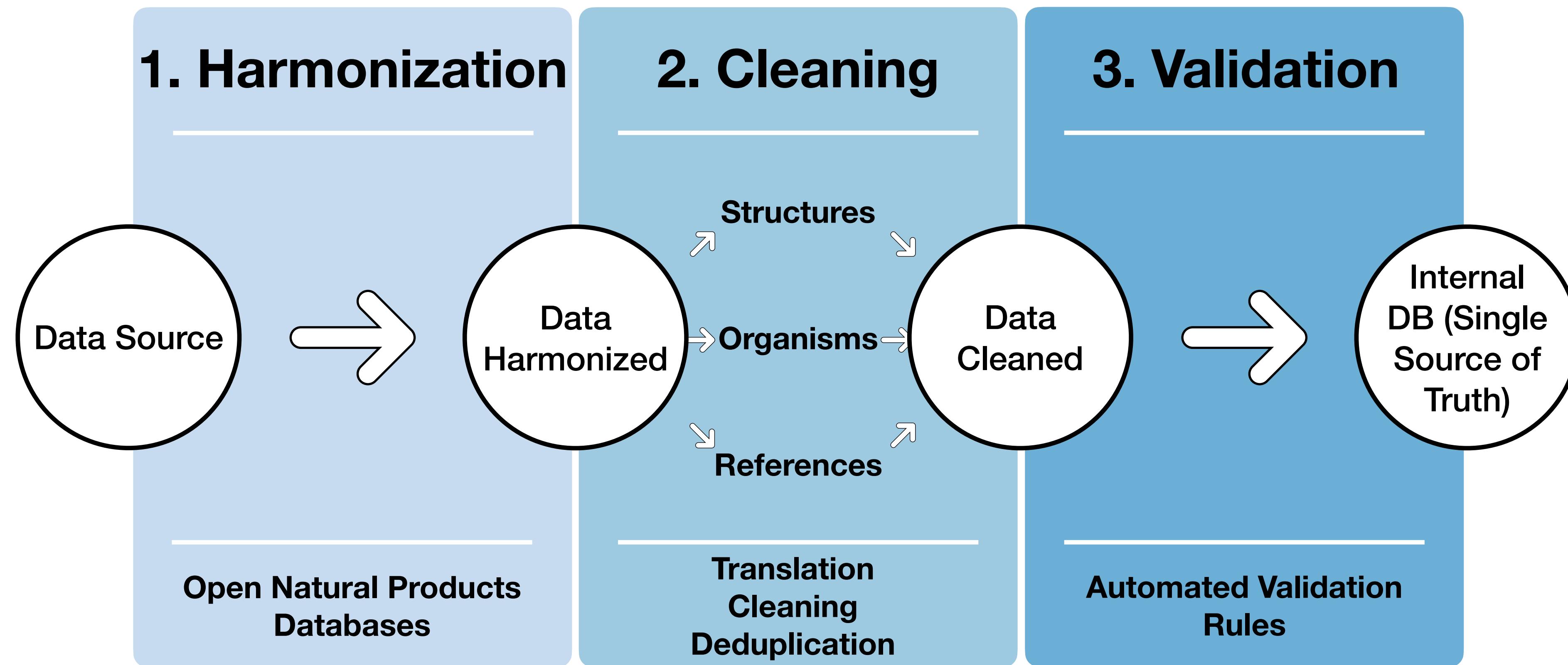
The initiative - *How?*



The initiative - How?

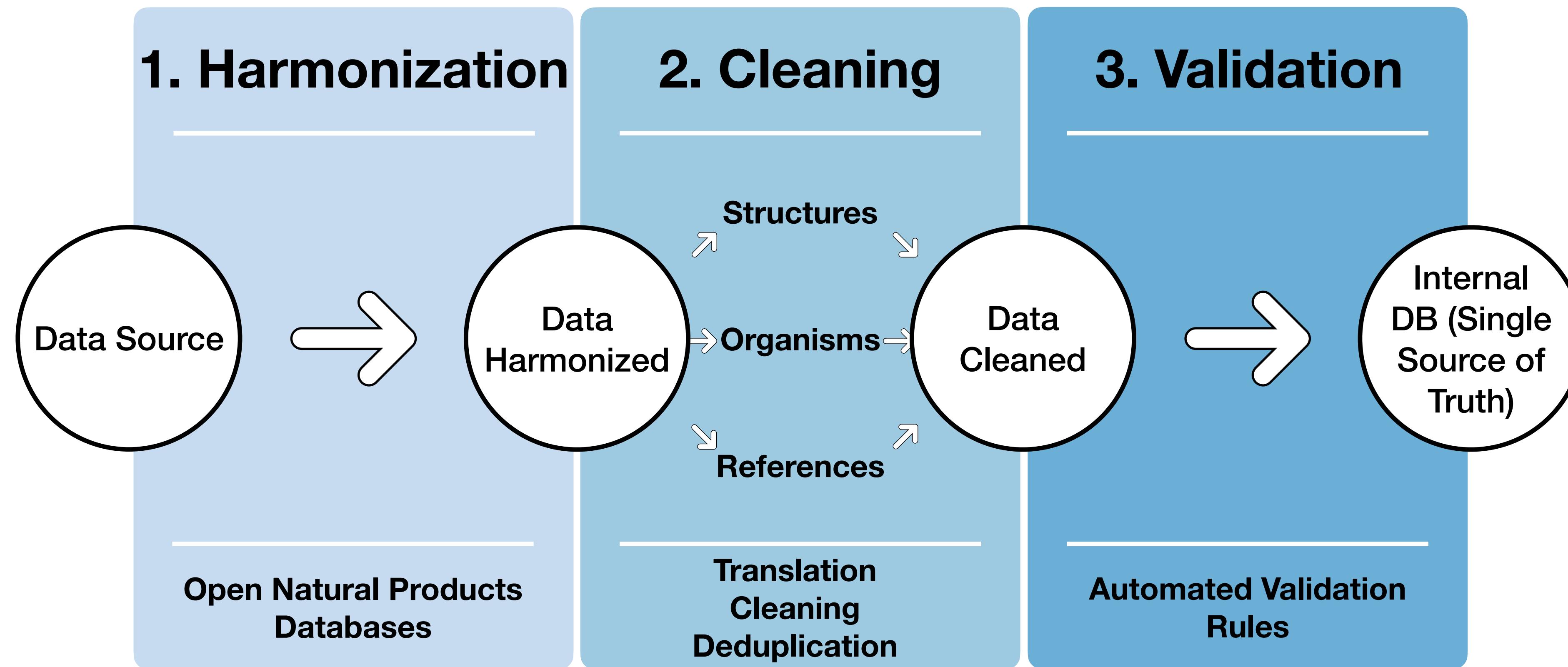


The initiative - How?



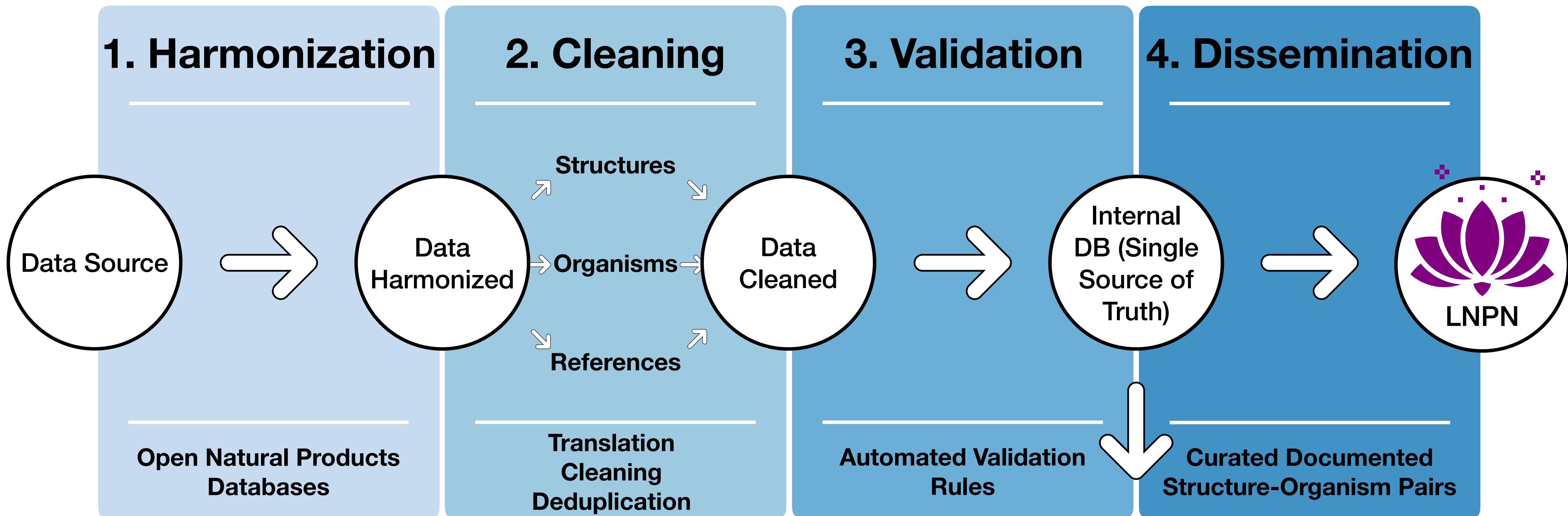
	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of <i>Cyathocalyx zeylanica</i> CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).

The initiative - How?



	Structure	Organism	Reference
Before curation	Cyathocaline	Stem bark of Cyathocalyx zeylanica CHAMP. ex HOOK. f. & THOMS. (Annonaceae)	Wijeratne E. M. K., de Silva L. B., Kikuchi T., Tezuka Y., Gunatilaka A. A. L., Kingston D. G. I., J. Nat. Prod., 58, 459-462 (1995).
After curation	VFIIVOHWCNHINZ-UHFFFAOYSA-N	<i>Cyathocalyx zeylanicus</i>	10.1021/NP50117A020

The initiative - *Where?*



The initiative - *Where?*



LOTUS

Find natural products
Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classific
[Structure Search](#) | [Advanced Search](#)

[Home](#) [Compound Browser](#) [Search](#) [Download](#) [Documentation](#)

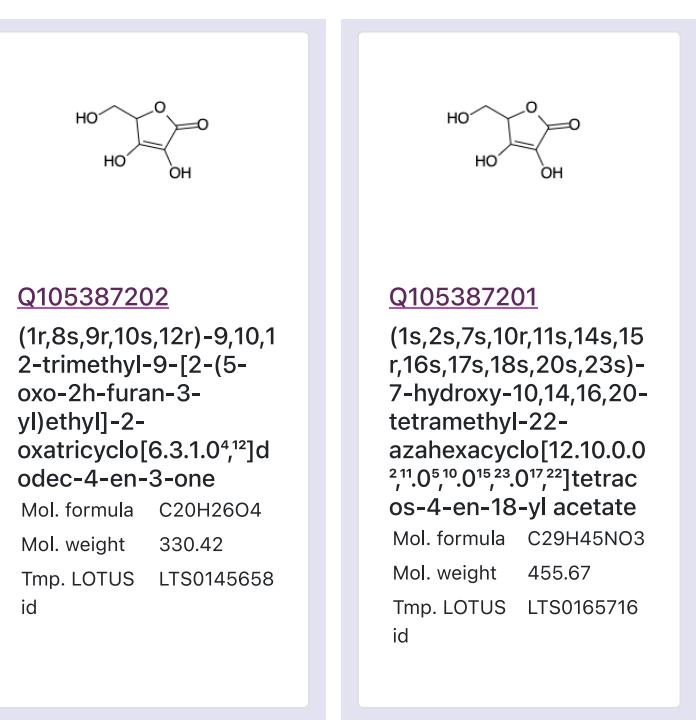
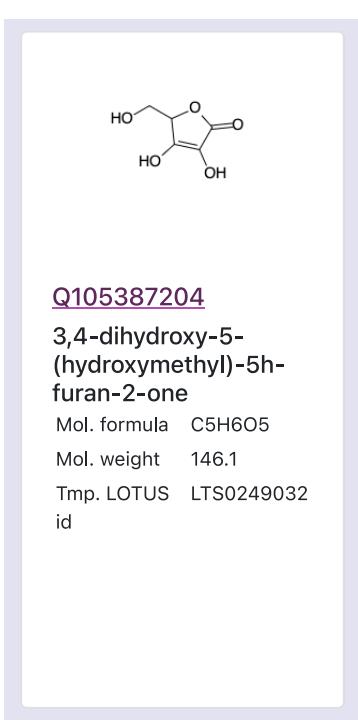
Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the naturalL prOducts occurrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

Component Browser

Cards Table

There are 276 518 natural products in the database

...



<https://lotus.naturalproducts.net/>

The initiative - *Where?*



Find natural products

Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classific

Search

Structure Search | Advanced Search

Home Compound Browser Search Download Documentation

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the naturalL prOducts occUrence database, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

Component Browser

Cards Table

There are 276 518 natural products in the database

« ⏴ 1 2 3 4 5 6 7 8 9 10 ... 11521 ⏵ »

[Q105387204](#)
3,4-dihydroxy-5-(hydroxymethyl)-5h-furan-2-one
Mol. formula C5H6O5
Mol. weight 146.1
Tmp. LOTUS LTS0249032
id

[Q27102265](#)
Lysopine
Mol. formula C9H18N2O4
Mol. weight 218.25
Tmp. LOTUS LTS0160430
id

[Q105387202](#)
(1s,2s,9r,10s,12r)-9,10,12-trimethyl-9-[2-(5-oxo-2h-furan-3-yl)ethyl]-2-oxatricyclo[6.3.1.0^{4,12}]dodec-4-en-3-one
Mol. formula C20H26O4
Mol. weight 330.42
Tmp. LOTUS LTS0145658
id

[Q105387201](#)
(1s,2s,7s,10r,11s,14s,15r,16s,17s,18s,20s,23s)-7-hydroxy-10,14,16,20-tetramethyl-22-azahexacyclo[6.3.1.0^{4,12}]tetracontane-4-en-18-yl acetate
Mol. formula C29H45NO3
Mol. weight 455.67
Tmp. LOTUS LTS0165716
id

<https://lotus.naturalproducts.net/>



Extraction

```
SELECT ?desired items
WHERE {
  VALUES ?classes {
    Chemical compound
  }
  ?item instance of ?classes.
  {
    ?item found in taxon ?stmt.
    ?stmt found in taxon ?taxon.
  }
  OPTIONAL {
    ?stmt occurrence ?ref.
    ?ref stated in ?art.
  }
}
SERVICE wikibase:label {
  language
  "[AUTO_LANGUAGE],en".
}
```

<https://www.wikidata.org/wiki/>

Wikidata:WikiProject Chemistry/Natural products

The initiative - *Wikidata*

Community

User can easily create an entry

User can curate previous entries

Easy query service



WIKIDATA

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

Item Discussion

Not logged in Talk Contributions Create account Log in

Read View history Search Wikidata

erysodine (Q27265641)

chemical compound

[edit](#)

Language	Label	Description	Also known as
English	erysodine	chemical compound	

Statements

instance of	chemical compound 0 references
mass	299.152±0 dalton 1 reference stated in PubChem PubChem CID 169017 language of work or name English title Erysodine (English) retrieved 14 October 2016

The initiative - *Wikidata*

isomeric SMILES	CO[C@H]1C[C@H]23C(=CCN2CCC4=CC(=C(C=C34)OC)O)C =C1
1 reference	
stated in	PubChem
PubChem CID	169017
language of work or name	English
title	Erysodine (English)
retrieved	14 October 2016

found in taxon	Erythrina edulis 1 reference stated in Alkaloid-Bearing Plants and Their Contained Alkaloids. 1957-1968
	Erythrina smithiana 1 reference stated in Alkaloid-Bearing Plants and Their Contained Alkaloids. 1957-1968

The initiative - SPARQL

« Hey Wiki, what are the compounds present in Mouse-ear cress (*Arabidopsis thaliana*)? »

The initiative - SPARQL

« Hey Wiki, what are the compounds present in Mouse-ear cress (*Arabidopsis thaliana*)? »

```
SELECT DISTINCT ?structure ?structureLabel ?inchiky
WHERE {
  ?structure wdt:P235 ?inchiky; # get the inchiky
  p:703 ?statement.           # « found in taxon » statement
  ?statement ps:703 wd:Q158695. # identifier of A. thaliana
```

Tip: you can remove the Qxxxxxx and hit Ctrl+space,
type the first letters of your organism and it should autocomplete

```
prov:wasDerivedFrom ?ref.      # get results with a reference
  ?ref pr:P248 ?art.          # get identifier of the reference
  ?art pr:P356 ?doi.          # get DOI of the reference
  SERVICE wikibase:label { language "[AUTO_LANGUAGE],en". }
}
```

The initiative - SPARQL

« Hey Wiki, what are the compounds present in Mouse-ear cress (*Arabidopsis thaliana*)? »

```
SELECT DISTINCT ?structure ?structureLabel ?inchiky  
WHERE {  
  ?structure wdt:P235 ?inchiky; # get the inchiky  
            p:703 ?statement.      # « found in taxon » statement  
  ?statement ps:703 wd:Q158695. # identifier of A. thaliana
```

Tip: you can remove the Qxxxxxx and hit Ctrl+space,
type the first letters of your organism and it should autocomplete

```
prov:wasDerivedFrom ?ref.      # get results with a reference  
  ?ref pr:P248 ?art.          # get identifier of the reference  
  ?art pr:P356 ?doi.          # get DOI of the reference  
SERVICE wikibase:label { language "[AUTO_LANGUAGE],en". }  
}
```

<https://w.wiki/3HMX>

The initiative - SPARQL

structure	structureLabel	structure_inchikey
Q (http://www.wikidata.org/entity/Q56229948) wd:Q56229948 (http://www.wikidata.org/entity/Q56229948)	lactose (stereochemistry undefined)	GUBGYTABKSRVRQ-QKKXWKRSA-N
Q (http://www.wikidata.org/entity/Q60998677) wd:Q60998677 (http://www.wikidata.org/entity/Q60998677)	L-pyroglutamic acid	ODHCTXKNWHHXJC-VKHMVHEASA-N
Q (http://www.wikidata.org/entity/Q61014519) wd:Q61014519 (http://www.wikidata.org/entity/Q61014519)	leucocyanidin (unspec. stereochem.)	SBZWTSHAFIOTE-UHFFFAOYSA-N
Q (http://www.wikidata.org/entity/Q57826492) wd:Q57826492 (http://www.wikidata.org/entity/Q57826492)	dehydroascorbic acid	SBJKKFFYIZUCET-UHFFFAOYSA-N
Q (http://www.wikidata.org/entity/Q104167739) wd:Q104167739 (http://www.wikidata.org/entity/Q104167739)	hexopyranosyl hexopyranoside	HDTRYLNUVZCQOY-UHFFFAOYSA-N
Q (http://www.wikidata.org/entity/Q29519883) wd:Q29519883 (http://www.wikidata.org/entity/Q29519883)	L-asparagine	DCXYFEDJOCDNAF-REOHCLBHSA-N
Q (http://www.wikidata.org/entity/Q104375437) wd:Q104375437 (http://www.wikidata.org/entity/Q104375437)	(2R,4R,5S,7S,12S,16S)-15-[(2S,3R,4R,5R)-3,4-dihydroxy-5,6-dimethylheptan-2-yl]-4,5-dihydroxy-2,16-dimethyl-9-oxatetracyclo[9.7.0.0 ^{2,7} .0 ^{12,16}]octadecan-8-one	IXVMHGVQKLDRKHFUFWTINHSA-N
Q (http://www.wikidata.org/entity/Q50362394) wd:Q50362394 (http://www.wikidata.org/entity/Q50362394)	5'-xanthyllic acid	DCTLYFZHFGENCW-UUOKFMHZSA-N
Q (http://www.wikidata.org/entity/Q104667554) wd:Q104667554 (http://www.wikidata.org/entity/Q104667554)	2-{6-[(4-hydroxy-3-methylbut-2-en-1-yl)amino]purin-9-yl}-6-(hydroxymethyl)oxane-3,4,5-triol	VYRAJOITMBSQSF-UHFFFAOYSA-N
Q (http://www.wikidata.org/entity/Q27139932) wd:Q27139932 (http://www.wikidata.org/entity/Q27139932)	β-L-idopyranose	WQZGKKKJIJFFOK-ZSNZIGRDSA-N
Q (http://www.wikidata.org/entity/Q27105114) wd:Q27105114 (http://www.wikidata.org/entity/Q27105114)	buddlenol B	LCXGTSCVCJANHX-AATRIKPKSA-N
Q (http://www.wikidata.org/entity/Q27105519) wd:Q27105519 (http://www.wikidata.org/entity/Q27105519)	4-methoxyglucobrassicin	IIAGSABLXRZUSE-UFRBAHOGSA-N
Q (http://www.wikidata.org/entity/Q27106307) wd:Q27106307 (http://www.wikidata.org/entity/Q27106307)	D-galactopyranose	WQZGKKKJIJFFOK-SVZMEOIVSA-N

912 compounds in a few milliseconds

The initiative - SPARQL

« Hey Wiki, which organisms are known to contain β -sitosterol? »

<https://w.wiki/3HLy>

The initiative - SPARQL

« Hey Wiki, which organisms are known to contain β -sitosterol? »

<https://w.wiki/3HLy>

« Hey Wiki, which organisms are known to contain stereoisomers of β -sitosterol? »

<https://w.wiki/3Jgs>

The initiative - SPARQL

« Hey Wiki, which organisms are known to contain β -sitosterol? »

<https://w.wiki/3HLy>

« Hey Wiki, which organisms are known to contain stereoisomers of β -sitosterol? »

<https://w.wiki/3Jgs>

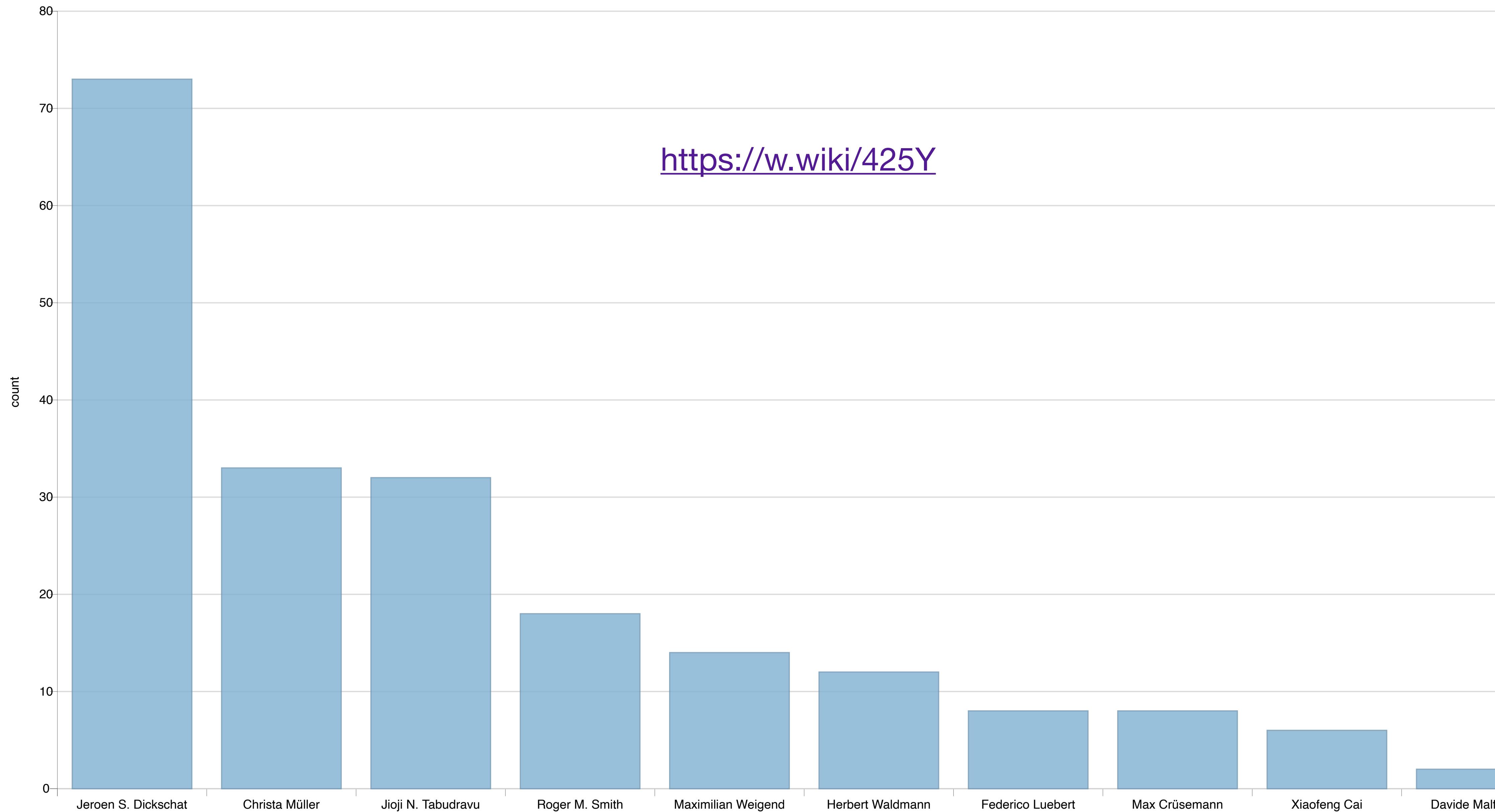
« Which pigments are found in which taxa, according to which reference? »

<https://w.wiki/3H3o>

The initiative - SPARQL

« Hey Wiki, can you count the number of structures found in taxon reported by researchers employed by the University of Bonn and rank them accordingly? »

The initiative - SPARQL



The initiative - SPARQL

« Hey Wiki, how many compounds are structurally similar to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

The initiative - SPARQL

« Hey Wiki, how many compounds are *structurally similar* to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

The initiative - SPARQL

« Hey Wiki, how many compounds are *structurally similar* to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

parent_taxon_name	count
Streptomyces	412
Streptomycetaceae	206
Apis	42
Penicillium	25
Torrubiella	25
Micromonospora	22
Saccharopolyspora	21
Kitasatospora	21
Fusarium	21
Albifimbria	21
Aspergillus	20

<https://w.wiki/3HMA>

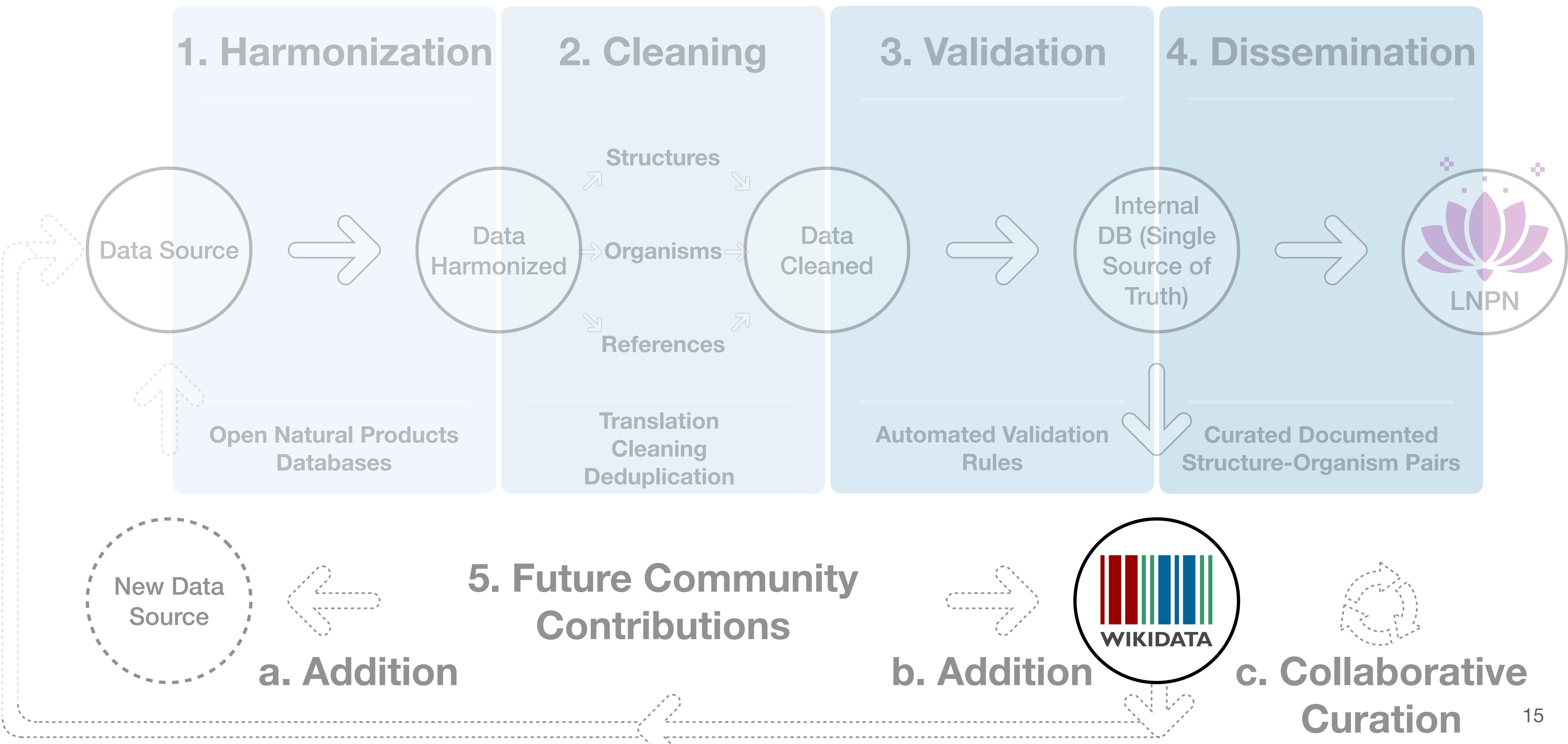
The initiative - SPARQL

« Hey Wiki, how many compounds are structurally similar to compounds labeled as antibiotics? (grouped by the parent taxon of the containing organism) »

parent_taxon_name	count
Streptomyces	412
Streptomycetaceae	206
Apis 	42
Penicillium	25
Torrubiella	25
Micromonospora	22
Saccharopolyspora	21
Kitasatospora	21
Fusarium	21
Albifimbria	21
Aspergillus	20

<https://w.wiki/3HMA>

The initiative - Perspectives



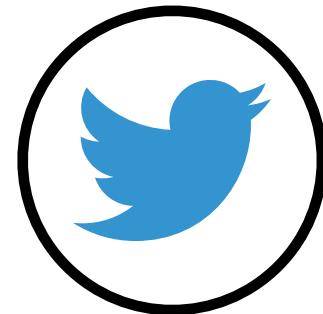
The initiative - Contact



<https://lotus.naturalproducts.net/>



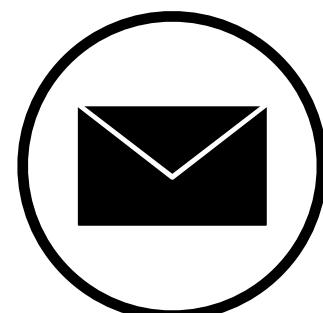
https://www.wikidata.org/wiki/Wikidata:WikiProject_Chemistry/Natural_products



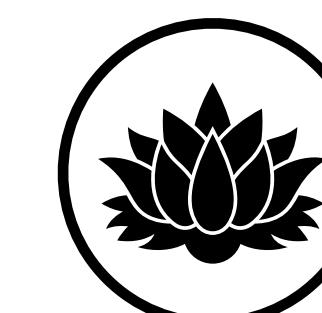
<https://twitter.com/TheLOTUSInitia1>



https://t.me/the_lotus_initiative



lotus-project@protonmail.com



<https://lotus.nprod.net/>