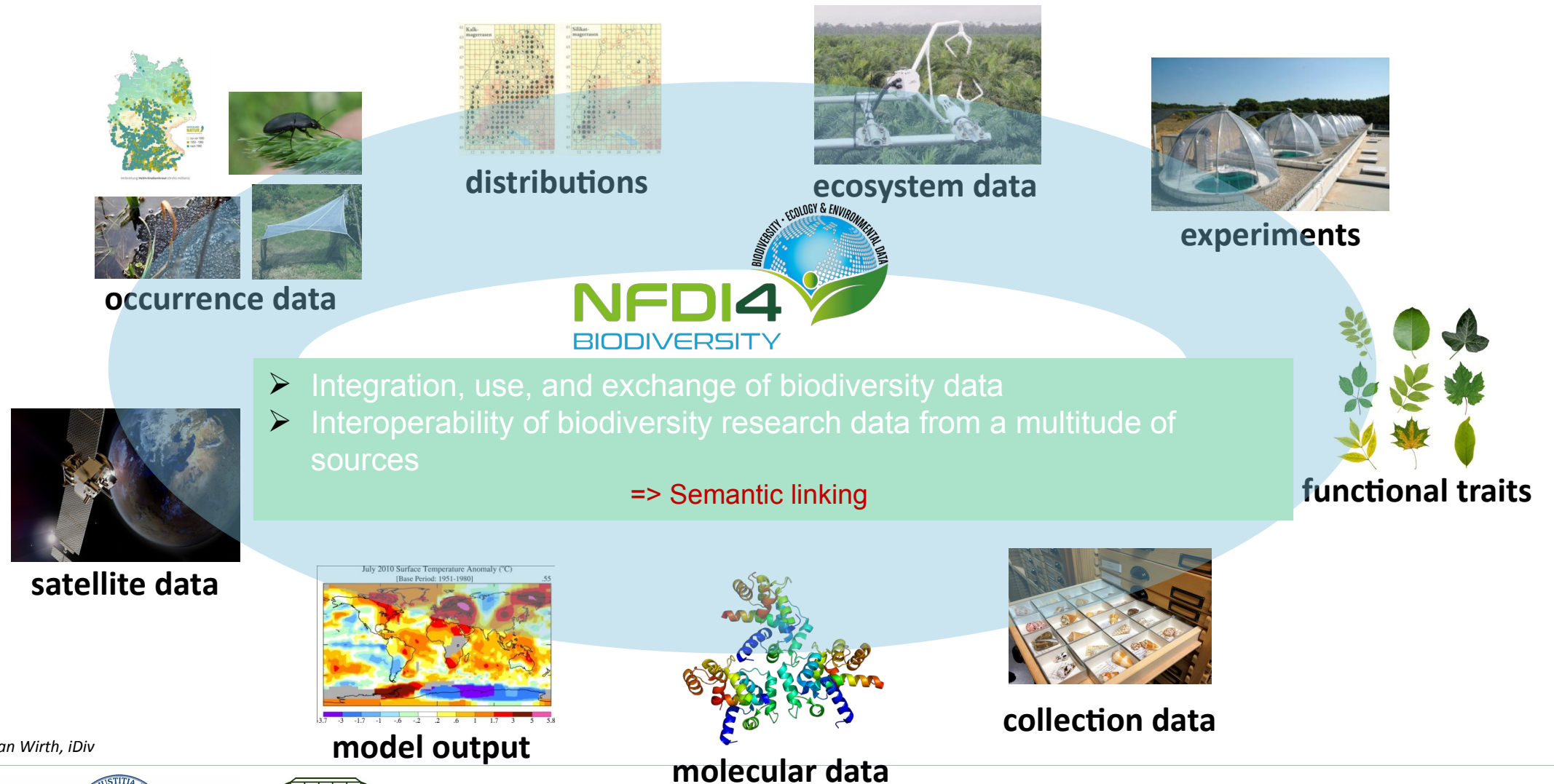


A pragmatic approach to concept-based annotation of scientific names in biodiversity and environmental research data

Andreas Kohlbecker, Anton Güntsch, Norbert Kilian, Wolf-Henning Kusber, Katja Luther, Andreas Müller, Eckhard von Raab-Straube & Walter Berendsohn

Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Germany, Department of
Biodiversity Informatics, Königin-Luise-Str.6-8, D-14195 Berlin

International data networks / NFDI4BioDiversity



Adapted from Christian Wirth, iDiv

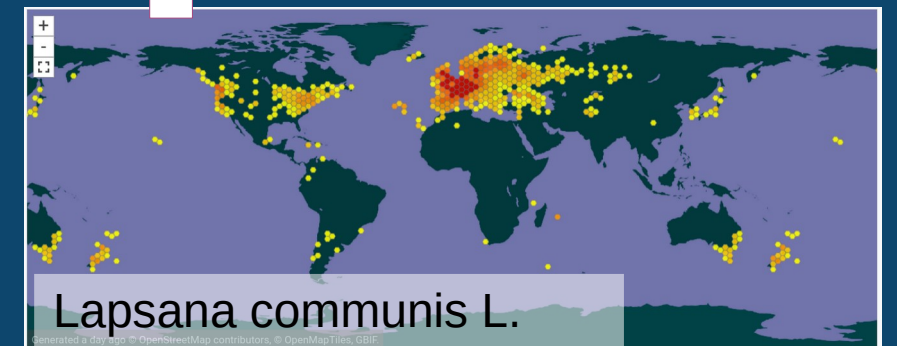
Biodiversity & environmental research data

- about or related to specific groups of organisms
- often ambiguously identified by the scientific name only
- A scientific name may be related to multiple taxon concepts!

==> Names are not enough!



Inference possible?



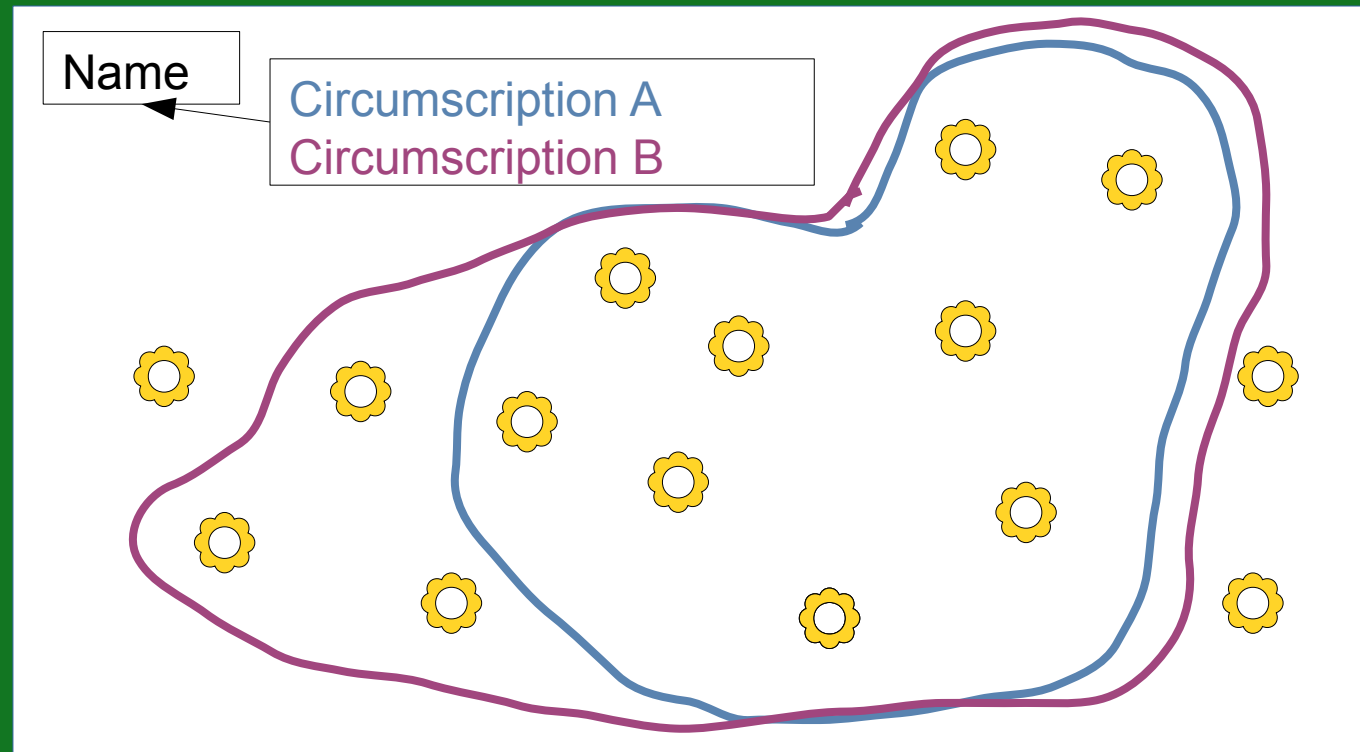
Interlude: Some biological basic knowledge

Why names are not enough: Taxon concepts

“Set of all individual organisms that fulfill all descriptive features covered by its circumscription and that are not fully covered by the circumscriptions of another taxon of the same rank.”

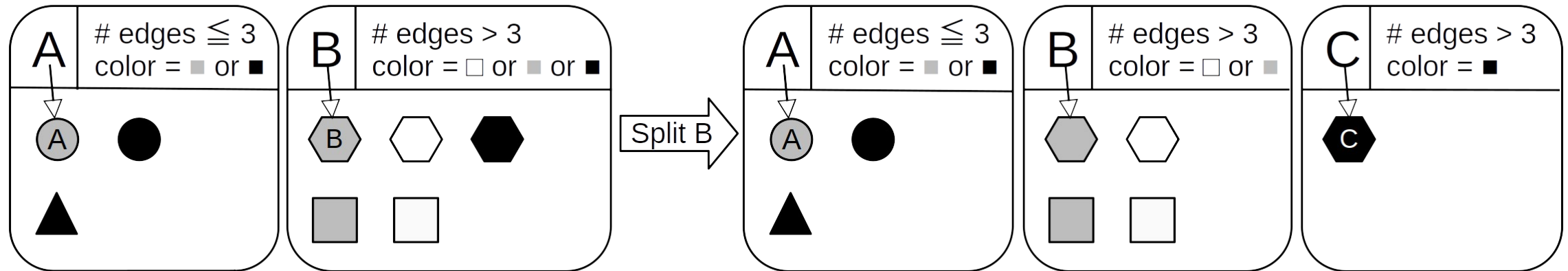
Criteria of taxon circumscriptions:

- Morphological & molecular characteristics
- Higher taxon concepts
- Type specimens
- ...



Taxonomic change

Taxon split operation



Uppercase single letters (A, B, C) are scientific names denoting taxon concepts (rounded boxes). Names are immutably bound to a single element per nomenclature (arrow). Each concept is defined by the circumscription consisting of the number (#) of edges and colors of the contained elements (colored shapes). The taxonomic operation “split” applied on “B” creates two new concepts named “B” and “C”. Even if bearing the same name “B”, the earlier and later taxon concepts are significantly different.

Circumscription / Type specimens

Scientific name: *Campanula formanekiana* Degen & Dörfli.

252. *Campanula Formanekiana* Deg. et Dörfli. (Taf. III, Fig. 7 a, b.)
Syn. *C. cinerea* Form. in Verb. d. naturf. Ver. in Brünn 1895, 1896, p. 37! non *Campanula cinerea* L. fil. Suppl. p. 139.
Macedonia centralis. In rupium fissuris prope Severni (floribus coeruleis); 24. Mai. (Exsicc. No. 237.)
Macedonia centralis. In rupium fissuris inter Allchar et Rožsdan (floribus albis); 21. Jun. (Exsicc. No. 236, utrisque sub n. *C. magnifica* nob. n. sp. l.)

Description

Descriptioni l. c. adde:
Herb. pub. calibus in rosulam congestis breviter incano tomentosis, in petiolum laminam longiorem, utrinque 8—10 denticulatum abrupte attenuatis; corollis e maximis (5—6 cm longis, 3—4 cm diam. tota superficie puberulis, intus glabris, ore non barbatis.

Stigmata 3—4; capsula trilocularis.

Formánek stellt diese prächtige Pflanze in die Nähe der *C. lanata* Friv. (*C. velutina* Vel. non al.); unserer Ansicht nach ist sie jedoch viel mehr mit der vom thessalischen Olymp und der Insel Euboea bekannten *C. incurva* Aucher ap. DC. Prodr. VII, p. 478 (*C. Leutweinii* Heldr. Appendix ad ind. sem. horti Athen. 1860, p. 7!) verwandt, von welcher sie sich durch die länger gestielten Rosetten-Blätter, die kürzeren Stengelblätter, die kleineren (5:3 mm) Kelchhängetel und die an der Mündung nicht gebärteten Glocken unterscheidet.

Es ist uns unbegreiflich, wie Boissier in der Flora orientalis III, p. 896 diese schon von De Candolle (a. a. O.) in die richtige Section gestellte und später von Heldreich a. a. O. noch ausführlicher und

First valid publication of the name = nomenclatural act

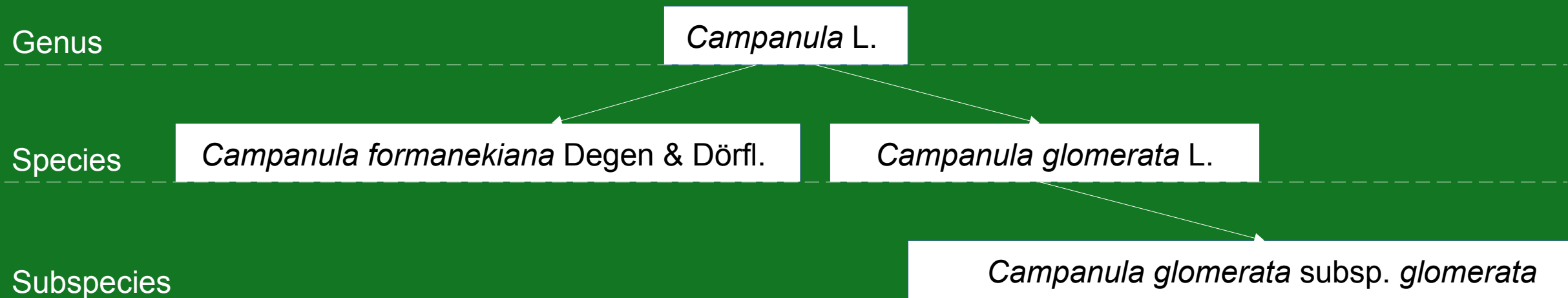


Type!

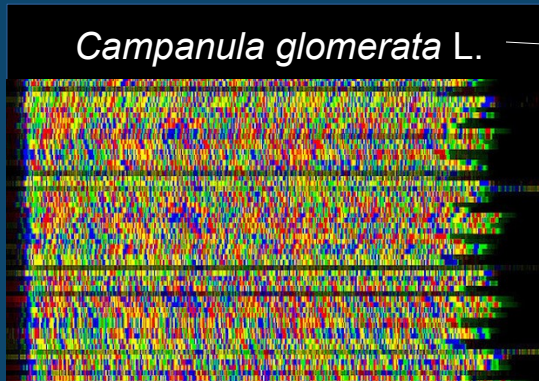
Circumscription / Taxonomic classification

Taxon concepts are hierarchically organized in classification trees

- Specific ranks = levels
- Directed non-cyclic graph
- Inheritance of taxon circumscription parent → child



Concept-based annotation of research data



Campanula glomerata L. **sec. Euro + Med (2021)**
PID= <http://euoplusmed/taxon/lQLTBJKVRcuc>

Campanula glomerata L. **sec. Govaerts, R. (1999). World Checklist of Seed Plants**
PID= <http://global.checklist.com/t/IRFa-VaY2b>

Campanula glomerata L. **sec. Wu, Z., Raven, P.H. & Hong, D. (eds.) (2011). Flora of China**
PID= <http://local.checklist.com/VaY2b6d95ccf5>

What is needed?

- Persistent identifiers (→ HTTP Identifiers)
- Definition of taxon concepts
- Reliable consistent workflow for taxon concept creation
- Matching of names in data to taxon concepts

Definition of taxon concepts

Schema to represent taxon concepts:

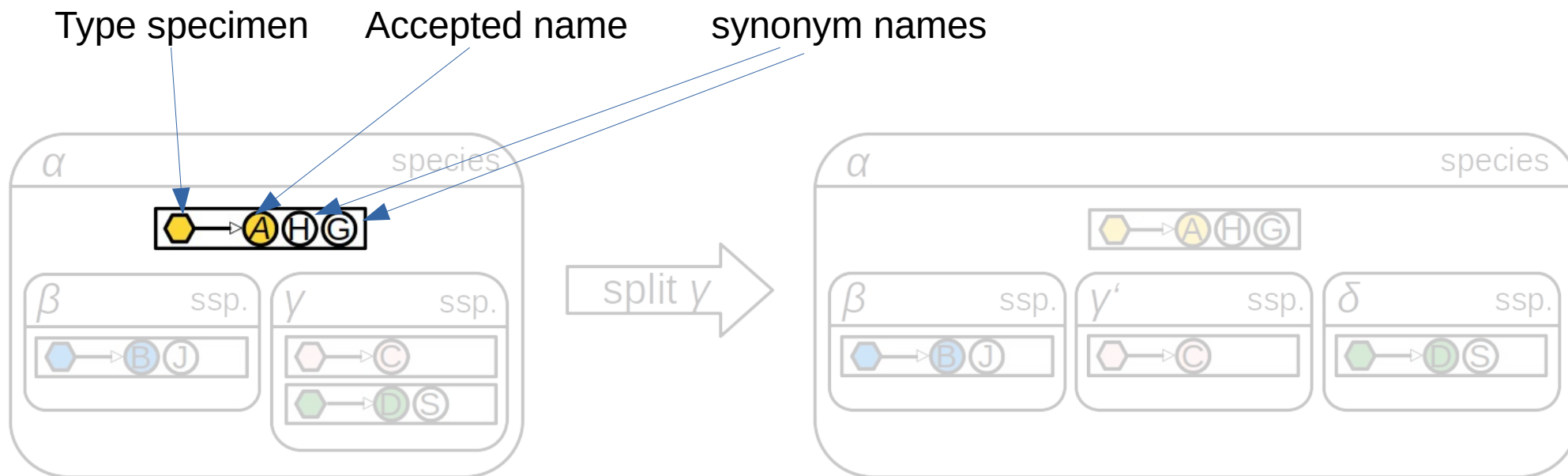
- **Linked Taxonomic Knowledge (LTK)** - Chawuthai, R. et al.: *Presenting and Preserving the Change in Taxonomic Knowledge for Linked Data*, *Semantic Web*. 7, pp. 589-616, 2016.
- **TAXREF** Michel, F. et al.: *A Model to Represent Nomenclatural and Taxonomic Information as Linked Data*, 2017
- **Taxon Concept Schema** Taxonomic Names and Concepts Interest Group: Taxonomic Concept Transfer Schema (TCS), version 1.01. Biodiversity Information Standards (TDWG), 2006
- ...

Missing:

Robust definition of taxon concepts (Needed to prevent taxa from being changed unnoticed)

A robust definition of taxon concepts

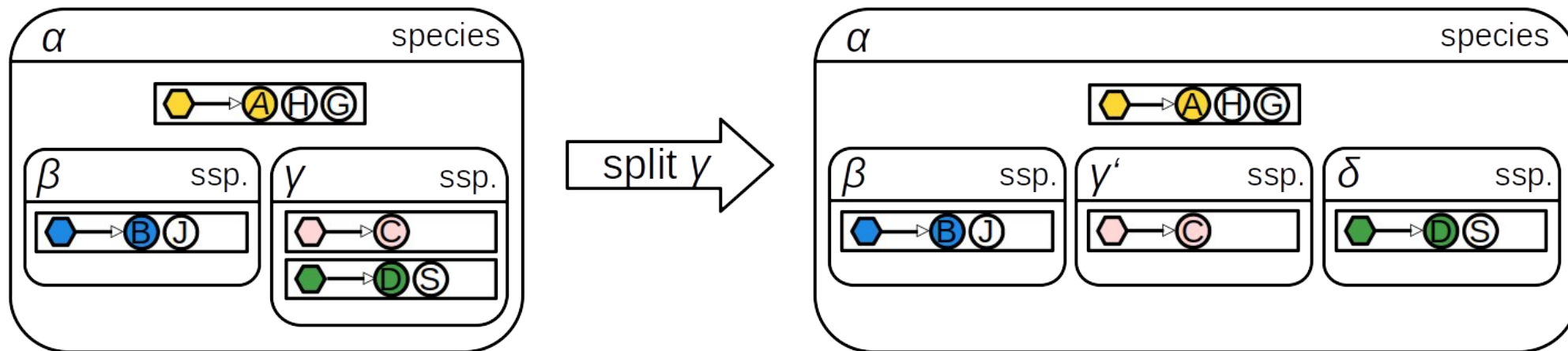
- Classifications, type specimens and synonymies to delimit taxonomic concepts.
- Relies on nomenclatural rules
- Avoids complexity of description based definitions



Taxon concept (rounded boxes) definition by classification and homotypic groups (rectangles) of **names (circle)** that are assigned to the same **type specimen (hexagon)**. The oldest name associated (arrow) with the type specimen is the accepted name, other names in the same group are synonyms.

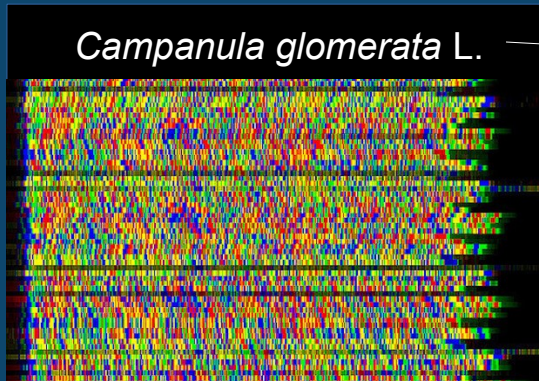
A robust definition of taxon concepts

- Classifications, type specimens and synonymies to delimit taxonomic concepts.
- Relies on nomenclatural rules
- Avoids complexity of description based definitions



Taxon concept (rounded boxes) definition by classification and homotypic groups (rectangles) of **names (circle)** that are assigned to the same **type specimen (hexagon)**. The oldest name associated (arrow) with the type specimen is the accepted name, other names in the same group are synonyms.

Concept-based annotation of research data



Campanula glomerata L. **sec. Euro + Med (2021)**
PID= <http://euoplusmed/taxon/lQLTBJKVRcuc>

Campanula glomerata L. **sec. Govaerts, R. (1999). World Checklist of Seed Plants**
PID= <http://global.checklist.com/t/IRFa-VaY2b>

Campanula glomerata L. **sec. Wu, Z., Raven, P.H. & Hong, D. (eds.) (2011). Flora of China**
PID= <http://local.checklist.com/VaY2b6d95ccf5>

What is needed?

- Persistent identifiers (→ HTTP Identifiers) [✓]
- Definition of taxon concepts [✓]
- Reliable consistent workflow for taxon concept creation []
- Matching of names in data to taxon concepts []

Traditional management of taxon concepts

- Manual management by editors
- Be aware of changes in the taxon circumscription
- Decision: taxonomic change or not?
- Create new taxon concept + persistent identifier

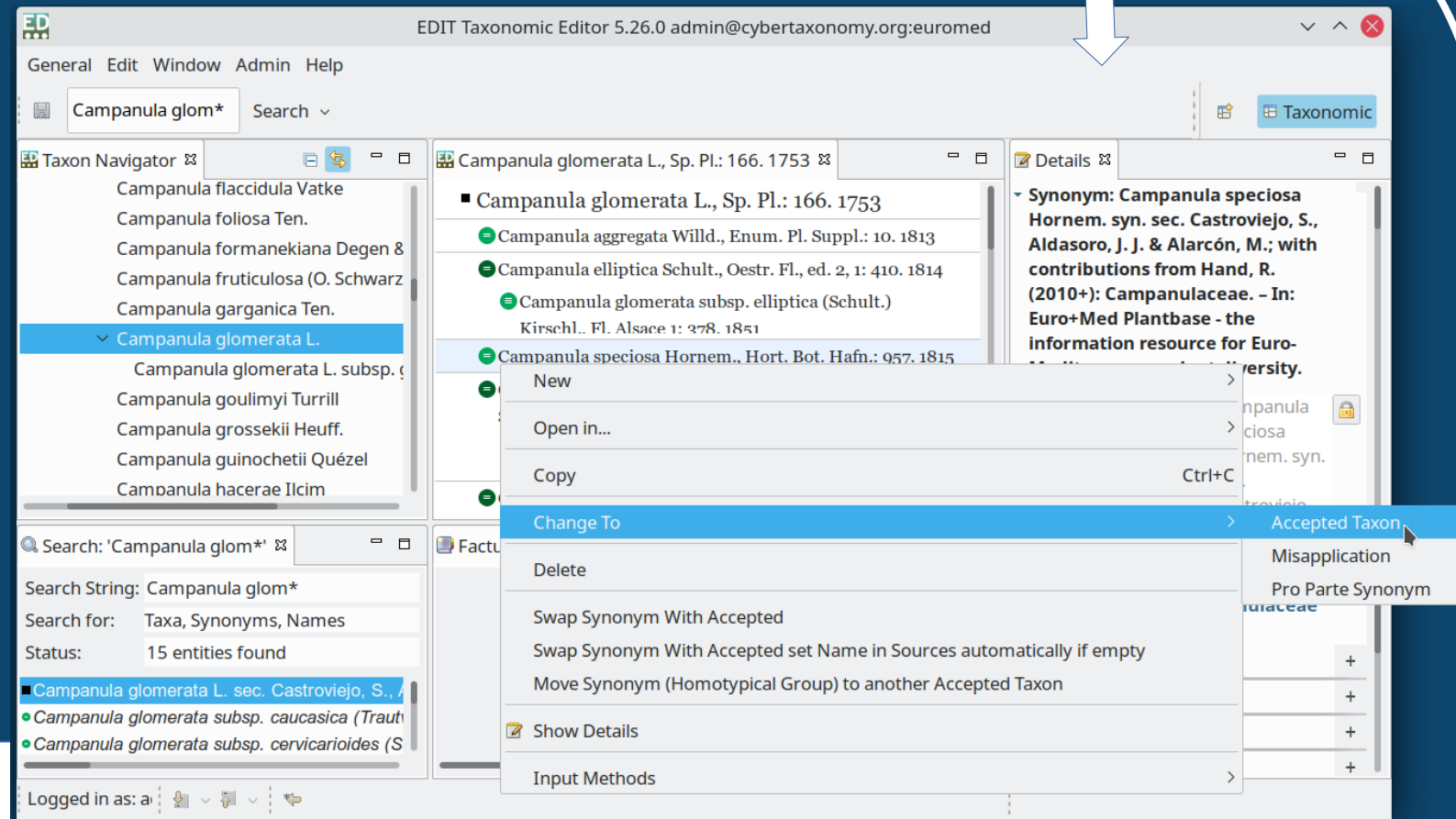
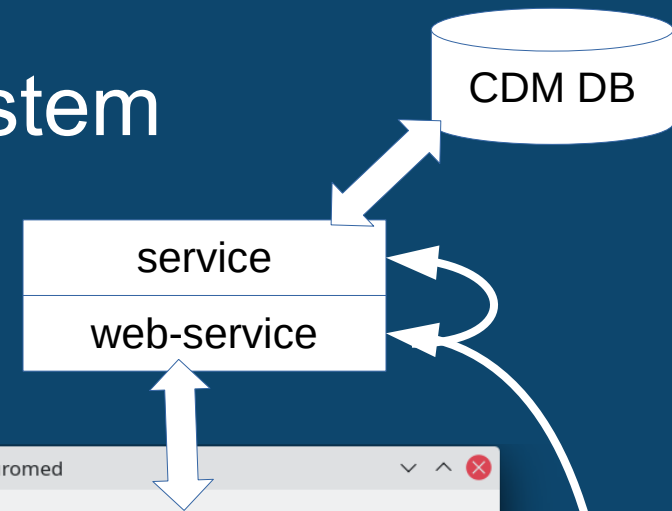
==> error prone!

Stabilizing taxon concepts by a rule-based system

Semi-automatic workflow actions at the transition points
= taxonomic operations that change the taxon circumscription

The 3 types of operations:

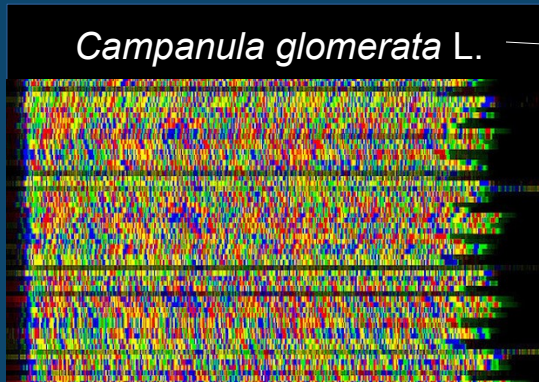
- **identity:**
no action
- **mandatory:**
automatic management in business layer (e.g. split operation)
- **optional:**
UI to ask editor for decision



Trigger split op.



Concept-based annotation of research data



Campanula glomerata L. **sec. Euro + Med (2021)**
PID= <http://euoplusmed/taxon/lQLTBJKVRcuc>

Campanula glomerata L. **sec. Govaerts, R. (1999). World Checklist of Seed Plants**
PID= <http://global.checklist.com/t/IRFa-VaY2b>

Campanula glomerata L. **sec. Wu, Z., Raven, P.H. & Hong, D. (eds.) (2011). Flora of China**
PID= <http://local.checklist.com/VaY2b6d95ccf5>

What is needed?

- Persistent identifiers (→ HTTP Identifiers) [✓]
- Definition of taxon concepts [✓]
- Reliable consistent workflow for taxon concept creation [✓]
- Matching of names in data to taxon concepts []

Matching of names

Services and web UIs to manually **match a scientific name the originally implicitly used taxon concept**

→ Meta lookup service covering all sources for taxon concepts in the data network

Potential complications:

- Multiple taxa exist for a name
- Originally implicitly used taxon cannot be found

Filtering criteria to facilitate

- Exact matching
- Decide that no provided taxon matches

Prototype implementation



NFDI 4
BIODIVERSITY
BIODIVERSITY, ECOLOGY & ENVIRONMENTAL DATA

Pilot project: Euro+Med PlantBase (<https://www.europusmed.org/>)



- Continuously updated by an international network of scientists
- Taxonomic reference system for
 - many regional floras and checklists
 - part of EU-Nomen (<http://www.eu-nomen.eu>)
--> EU INSPIRE directive

Thank you!

Questions ?

a.kohlbecker@bgbm.org

<http://www.bgbm.org/BioDivInf/>

<https://cybertaxonomy.eu/>

<https://www.nfdi4biodiversity.org/de/>