How to Search for Biological Data?

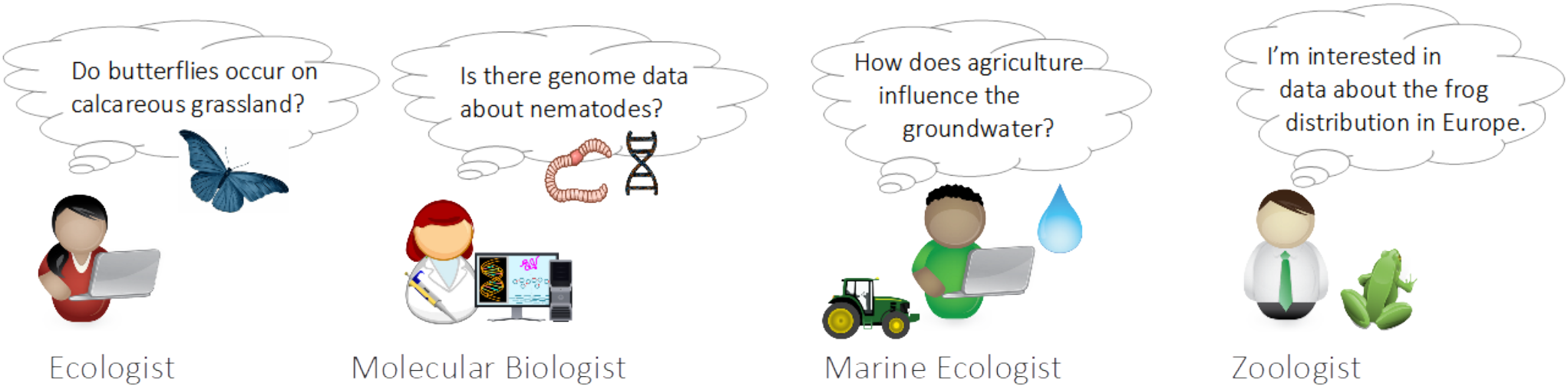
A Comparision of User Interfaces in a Semantic Data Search

Felicitas Löffler[1], Friederike Klan[2], Birgitta König-Ries[1]

[1]Heinz Nixdorf Chair for Distributed Information Systems
Department of Mathematics and Computer Science
Friedrich Schiller University Jena, Germany

[2]German Aerospace Center (DLR) , Jena, Germany

MOTIVATION



Scholars have complex information needs

Metadata Standards

EML

```
<eml><title>Tree measurements in a
subtropical forest
</title>
<creator>Hans Meier</creator>
<description> We analyzed […] </description>
<keywords>schima superba</keywords>
<methods></methods>
…
</eml>
```

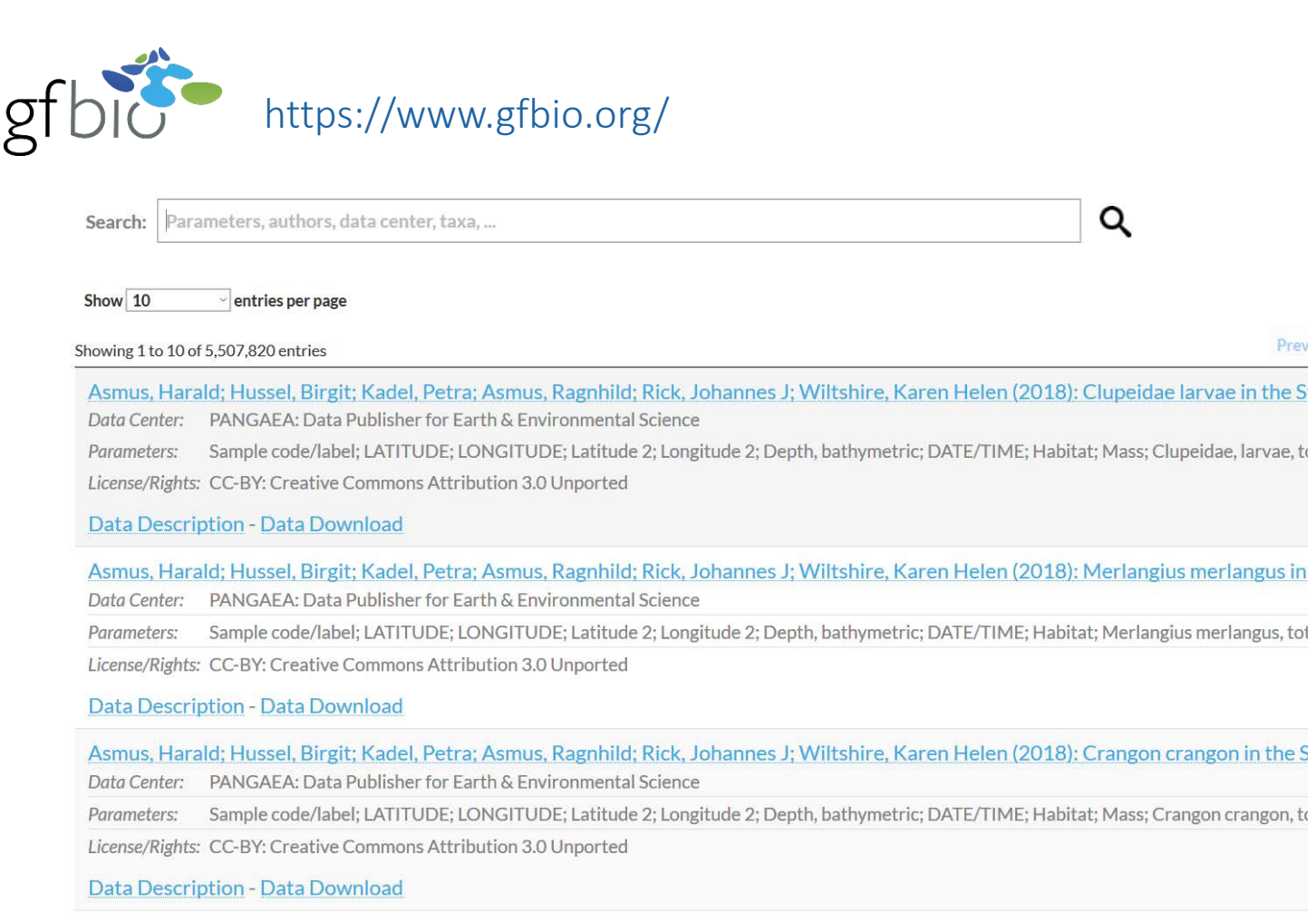
ABCD

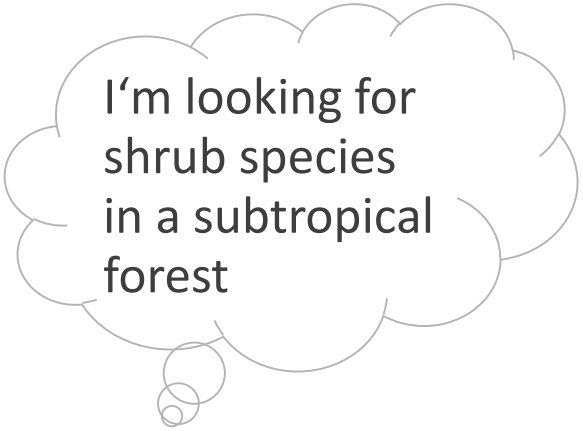
```
<abcd:dataset>
<abcd:TechnicalContact>Museum
XYZ</abcd:TechnicalContact>
<abcd:title>Herbarium </abcd:title>
…
<abcd:ScientificName>Helianthus
annus</abcd:ScientificName>
…
</abcd:dataset>
```

What? Who? Where? When?

→ Metadata are focused on data and citation aspects

https://www.gfbio.org/





search over organisms, processes or habitats is not possible

How to search for biological data?

1.) What information categories are required to determine the search context?

2.) How to enable data search over these categories?

3.) What kind of user interface would be appropriate considering the categories?

OUR APPROACH

Information Categories

A) Question Corpus with 169 search and research questions, e.g.:

Do butterflies occur on calcareous grassland?

What are associated taxa?

Felicitas Löffler, felicitas.loeffler@uni-jena.de
Heinz Nixdorf Chair for Distributed Information Systems

Information Categories

A) Question Corpus with 169 search and research questions, e.g.:

Do butterflies occur on calcareous grassland?

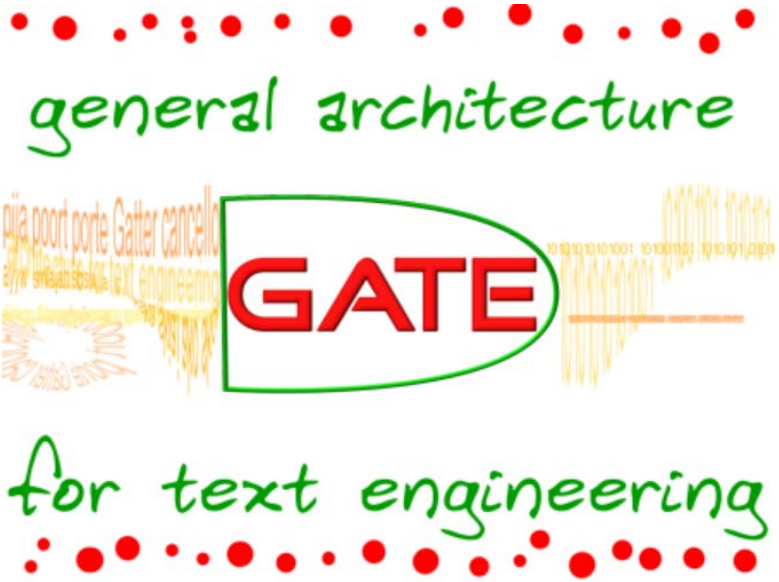
What are associated taxa?

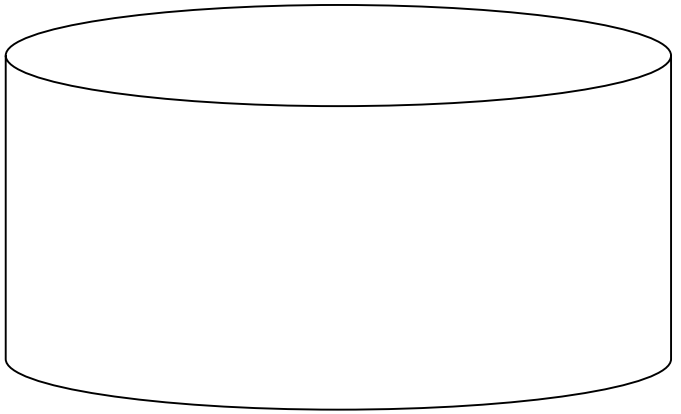
B) 13 categories were identified, evaluation with 9 domain experts, finally 10 categories have been verified

Organism, Environment, Process, Material, Quality, Event, Person, Location, Time, Human Intervention

Felicitas Löffler, felicitas.loeffler@uni-jena.de
Heinz Nixdorf Chair for Distributed Information Systems

Categories – Ontologies





Organism	Environment	Quality	Process	Material
OrganismTagger	ENVO (parts)	PATO, TO, FLOPO	ENVO (parts), GO (parts), UEBERON (parts)	ENVO (parts), ChemicalTagger

USER INTERFACES

User Interface I

- form-based - one search field per category
- SPARQL templates per category
- includes synonyms and more specific results

User Interface II

- NLP based - category is determined automatically
- SPARQL template over all categories for a first result
- includes synonyms and more specific results

Felicitas Löffler, felicitas.loeffler@uni-jena.de
Heinz Nixdorf Chair for Distributed Information Systems

Demo

User Interface I: https://cloud.uni-jena.de/index.php/s/iibAyfwPcmGMjLo

User Interface II: https://cloud.uni-jena.de/index.php/s/mcCTiHqF8ZHy7Sj

Felicitas Löffler, felicitas.loeffler@uni-jena.de
Heinz Nixdorf Chair for Distributed Information Systems

Which interface do you prefer?

Please take part in a short interactive survey!

www.mentimeter.com

Code: 775281

Thank You! Questions?

@felicitasloe

@fusionUniJena