

# Earth System's Biology Data FAIRification: Mobilizing SGN Collection Data for the BITS Terminology Service in Earth System Science

*Iriomote cat illustration by Prof. Mitsuru Moriguchi, Okinawa University.*



## Why we do what we do?

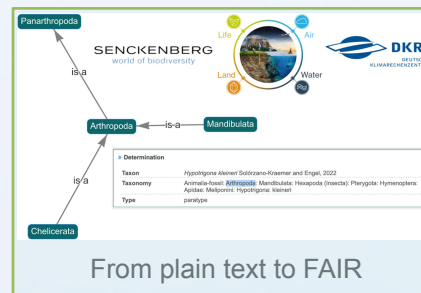
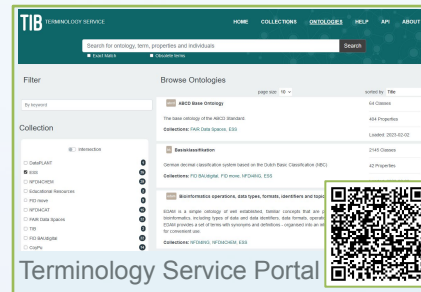
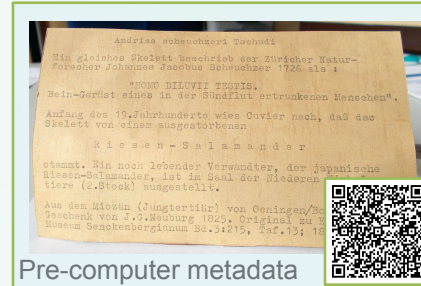
- Institution with 11 facilities
- over 40 Million exhibition and research items
- Very heterogeneous data and data infrastructures:

1. Common database: <https://search.senckenberg.de> currently 1.521.698 objects in 124 collections
2. Facility and department maintained implementations
3. Locally stored data

Conclusion: Only a partial FAIRness is given with a potentially wide range of existing and upcoming items and collections.

## BITS Terminology Service Overview

- Terminologies, specifically for RDF and OWL
- Access for humans and machines
- Curation of new terms (vocabulary)
- Metadata management
- Visualization of semantic artifacts



## Preparation and integration

1. Process and identify relevant data fields
2. Identify relevant elements such as words and phrases
  - a. safely remove irrelevant content such as numbers
  - b. remove redundancy to improve performance
3. Use BITS TS to obtain metadata about the elements
4. Process unresolvable phrases by splitting them into words
5. **Selecting the best results**
6. Store the resulting data set in an appropriate manner for human and machine agents (in progress)

## Results and lessons learned

- Using pre-FAIR data sets requires a lot of manual work
- Identifying “the best result” (of over 30.000) for each element is not possible within an appropriate time frame: Using **coverage based approach** for the whole data set

## Conclusion

Based on GO FAIR, we are already following most of the principles using the SGN infrastructure and BITS TS. Full implementation of the BITS TS extension to the SGN repository will provide a long-term strategic approach to providing machine-actionable digital specimens and enhancing FAIRness.