

Biodiversity_next pre-conference: Biodiversity Informatics 101

3 steps to use Semantic Web technologies

NAKAZATO, Takeru
@tkr_nak

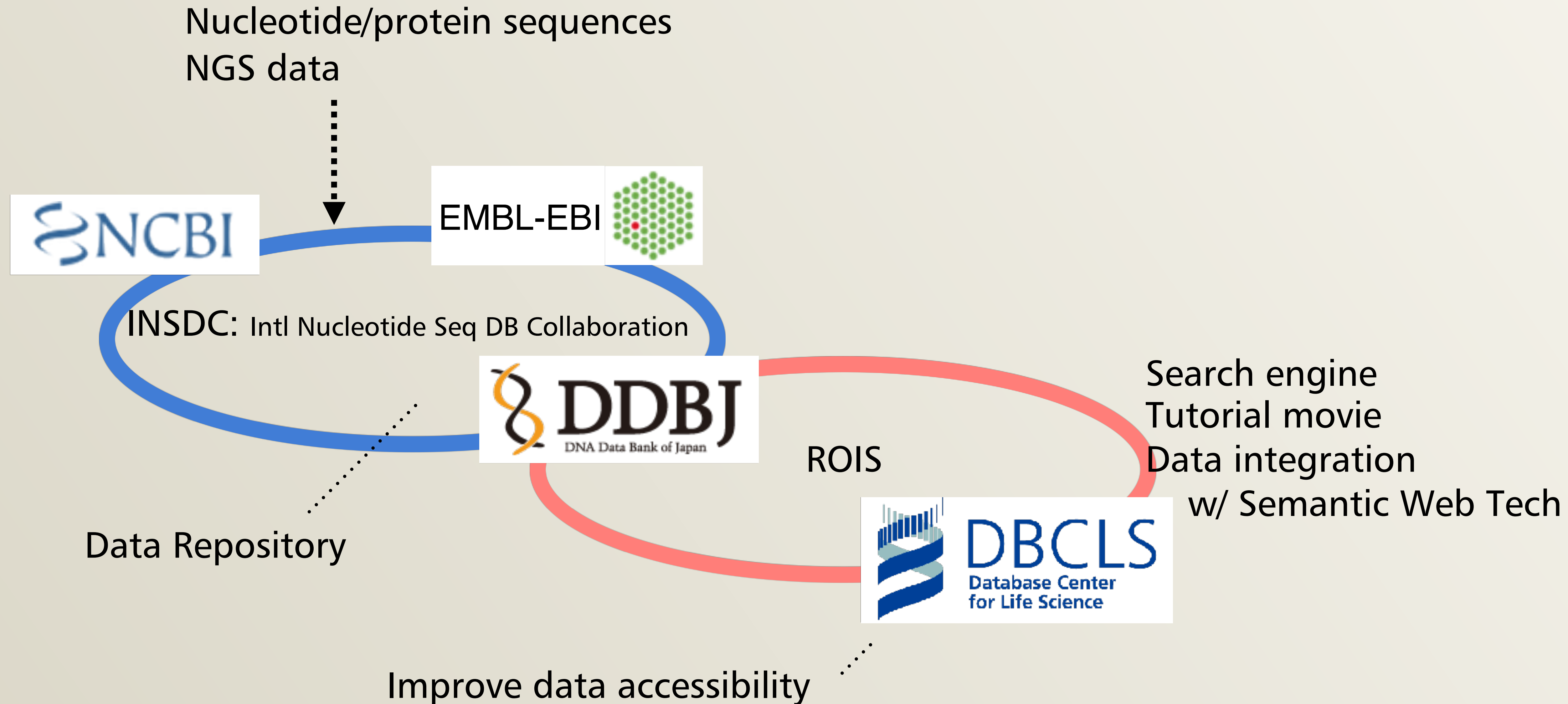


Database Center for Life Science (DBCLS),
Joint Support-Center for Data Science Research,
Research Organization of Information and Systems (ROIS), JAPAN



2019/10/21
@Scheltema, Leiden, the Netherlands

Public data for sequence and our mission



Semantic Web technology

RDF

Ontology

Linked Open Data

JFYI: RDF (Resource Description Framework)



Old world swallowtail



eats

fennel .



<https://eol.org/media/7780539>
CC-BY-SA

N-triples format

```
<http://taxon.db/Papilio_machaon> <http://relation.org/eats> <http://taxon.db/Foeniculum_vulgare> .
```

Turtle format


```
@prefix taxon: <http://taxon.db/>  
@prefix rel: <http://relation.org/>
```

```
taxon:Papilio_machaon rel:eats taxon:Foeniculum_vulgare .
```

* These URIs are dummy

Ontology

EMBL-EBI




Ontology Lookup Service



[Home](#)
[Ontologies](#)
[Documentation](#)
[About](#)



[OLS](#) > [NCBI organismal classification](#)


[NCBITAXON](#) > [NCBITaxon:76193](#)



Papilio machaon


http://purl.obolibrary.org/obo/NCBITaxon_76193


 Tree view
  Term history

 Graph view

Reset tree

Show all siblings

- root
 - cellular organisms
 - Eukaryota
 - Opisthokonta
 - Metazoa
 - Eumetazoa
 - Bilateria
 - Protostomia
 - Ecdysozoa
 - Panarthropoda
 - Arthropoda
 - Mandibulata
 - Pancrustacea
 - Hexapoda

Term info

[database cross reference](#)

- GC_ID:1

[common name](#)

- Old World swallowtail

[genbank common name](#)

- common yellow swallowtail

[common name](#)

- artemisia swallowtail

[has exact synonym](#)

- common yellow swallowtail, Old World swallowtail
- artemisia swallowtail

[has obo namespace](#)

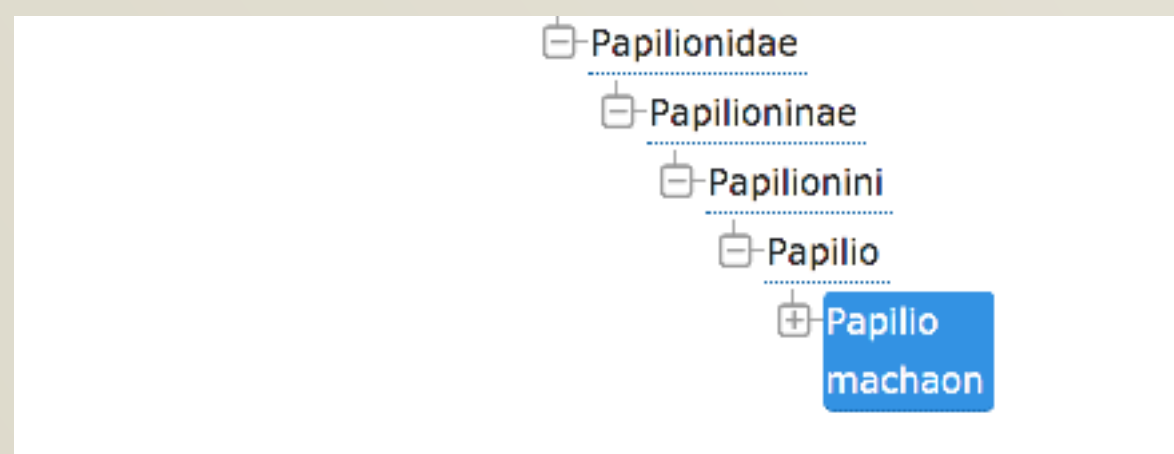
- ncbi_taxonomy

- Controlled vocabulary thesaurus or keyword collection as a reference
- hierarchically structured
- The Relationship between parent and child of term is defined such as "part of", "is a", and "subclass of".

<http://taxon.db/Papilio_machaon>

Ontology

Data value




Example: EOL

<http://taxon.db/Papilio_machaon> <http://relation.org/eats> <http://taxon.db/Foeniculum_vulgare> .

eol.org/pages/130925/data

eol en

Life » ... » Animals » ... » Arthropods » ... » Hexapods » Insects » Winged insects » ... » Insects with



Asian Swallowtail

Papilio xuthus Linnaeus 1767

overview

data

media

articles

maps

names

filter by attribute

filter by provider

eats

Toddalia asiatica (L.) Lam.

URI: http://purl.obolibrary.org/obo/RO_0002470

show all records

host records of butterfly

eats

Zanthoxylum armatum subtrifoliatum

eats

Tetradium ruticarpum (A. Juss.) T. G. Hartley

eats

Ontobee

Home

Intro

Statistics

SPARQL

Ontobee

Annotator

Tutorial

FAQs

Relation Ontology

Keywords:

Search terms

ObjectProperty: eats

Term IRI: http://purl.obolibrary.org/obo/RO_0002470

Annotations

definition editor: Chris Mungall

alternative term: is subject of eating interaction with

in_subset: eco subset

Property Hierarchy

topObjectProperty

+ [ecologically related to](#)

+ [biotically interacts with](#)

+ [participates in a biotic-biotic interaction with](#)

+ [trophically interacts with](#)

- [preys on](#)

- [acquires nutrients from](#)

- [preyed upon by](#)

- [provides nutrients for](#)

- [is eaten by](#)

- [eats](#)

Superproperties

+ [trophically interacts with](#)

JFYI: SPARQL as RDF Query Language

[eol.db/relation/graph/]

```
@prefix taxon: <http://taxon.db/>  
@prefix rel: <http://relation.org/>
```

```
taxon:Papilio_machaon rel:eats taxon:Foeniculum_vulgare .
```

[SPARQL example]

```
PREFIX taxon: <http://taxon.db/>  
PREFIX rel: <http://relation.org/>  
  
SELECT ?plant  
FROM <http://eol.db/relation/graph/>  
WHERE {  
    taxon:Papilio_machaon rel:eats ?plant.  
}
```



Foeniculum_vulgare

Traditional relation database and SQL

| taxon | Rel | plant |
|-----------------|------|--------------------|
| Papilio_machaon | eats | Foeniculum_vulgare |
| Papilio_xuthus | Eats | Citrus aurantium |
| | ... | |

[SQL example]

```
SELECT plant from EOL_REL  
where taxon="Papilio_machaon" and rel="eats";
```



Foeniculum_vulgare

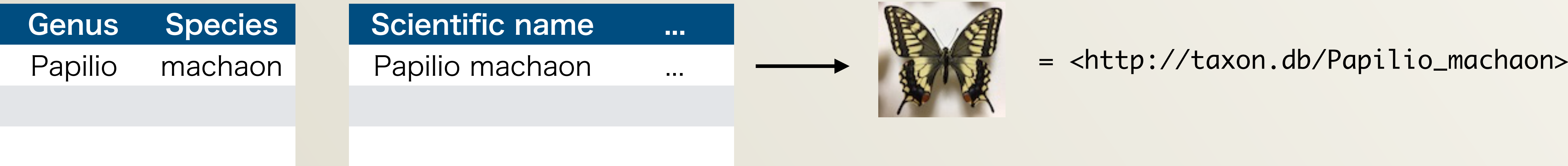
Oh, relational database is also difficult... But I can use it in Excel...

What makes RDF unfamiliar is due to technical aspects.

- How to describe RDF?**
- How to store RDF data?**
- How to use RDF data?**

The concept of Semantic Web technology

- Standardization: Common keys and common value



- Reusable: No need to download and convert someone's DB to local data

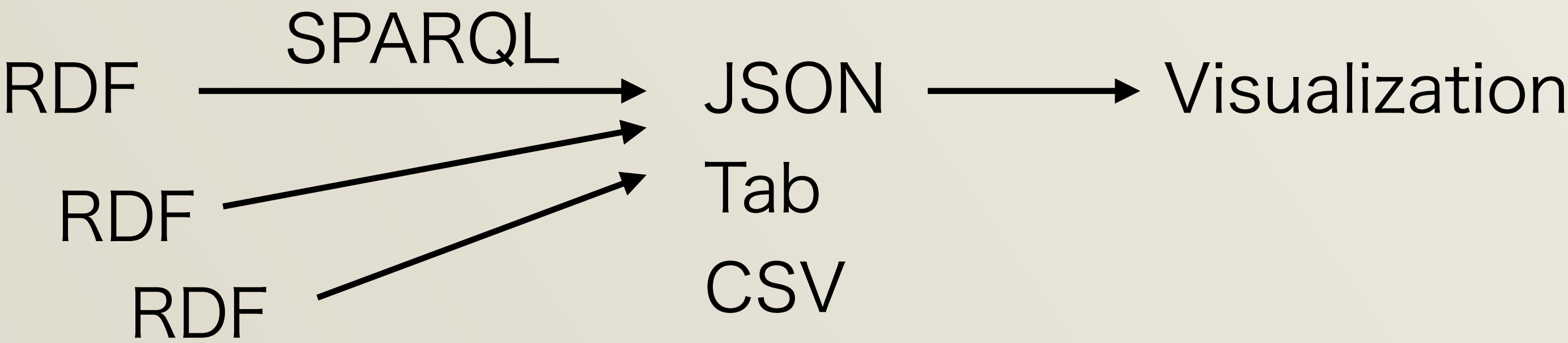
```
@prefix taxon: <http://taxon.db/>
@prefix rel: <http://relation.org/>

taxon:Papilio_machaon rel:eats taxon:Foeniculum_vulgare .
```

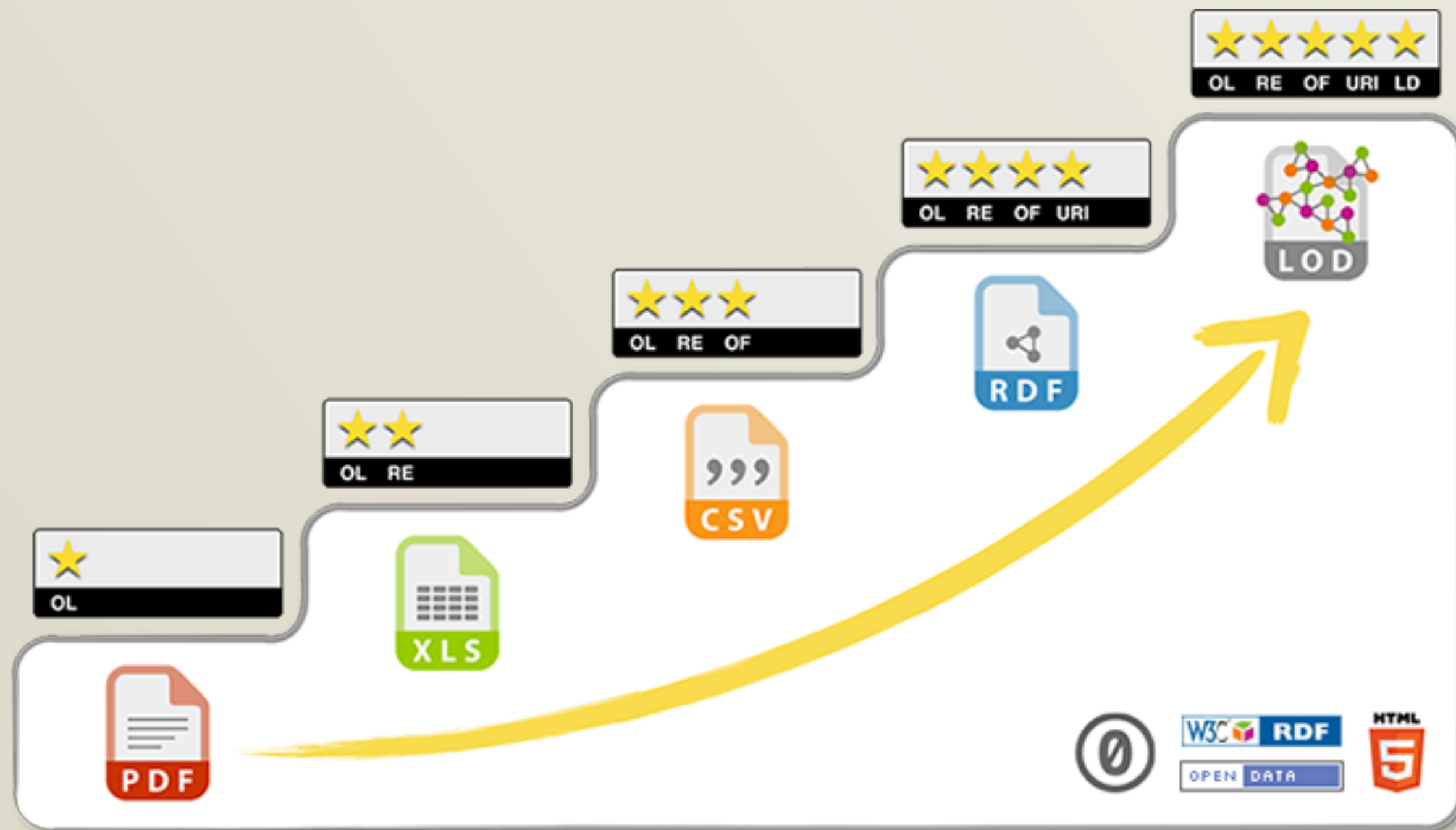
Someone already provide taxonomy DB

So only call DB with value

- Usablity: Easy to parse and easy to merge



5★ OPEN DATA



- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)³
- ★★★★ use URIs to denote things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data.

<https://5stardata.info/>

Step 1: Semantic Web tech with Excel

| | A | B | C | D | E | F | |
|---|-------------------------------|--|-------------|--------------|---------|-----------------|-----|
| 1 | scientificName | taxonID | order | family | genus | specificEpithet | col |
| 2 | Papilio machaon Linnaeus 1758 | http://purl.obolibrary.org/obo/NCBITaxon_76193 | Lepidoptera | Papilionidae | Papilio | machaon | NH |
| 3 | Papilio xuthus Linnaeus 1767 | | Lepidoptera | Papilionidae | Papilio | xuthus | NH |

Data value:
Follow Darwin Core's manual
or use the value in major databases.
Use controlled vocabularies as
possible.

Data value:
URI is better, but no required

```
<http://local.db/submission/123> dwc:taxonID taxon:NCBITaxon_76193 .  
taxon:NCBITaxon_76193 rdfs:label "Papilio machaon" .  
...
```


Step 2: Try public SPARQL Endpoint

← → ↺

保護されていない通信 | togogenome.org/sparql

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX taxon: <http://ddbj.nig.ac.jp/ontologies/taxonomy/>

SELECT ?taxid ?label ?rank
FROM <http://togogenome.org/graph/taxonomy>
WHERE {
 ?taxid_self rdfs:label "Papilio machaon" .
 ?taxid_self rdfs:subClassOf* ?taxid .
 ?taxid rdfs:label ?label ;
 taxon:rank ?rank .
 VALUES ?rank {
 taxon:Class
 taxon:Order
 taxon:Family
 taxon:Genus
 taxon:Species
 }
}

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML

Execution timeout: 0 milliseconds (values less than 1000 are ignored)

Options: ☒ Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

| taxid | label | rank |
|---|-------------------|---|
| http://identifiers.org/taxonomy/7088 | "Lepidoptera" | http://ddbj.nig.ac.jp/ontologies/taxonomy/Order |
| http://identifiers.org/taxonomy/7143 | "Papilionidae" | http://ddbj.nig.ac.jp/ontologies/taxonomy/Family |
| http://identifiers.org/taxonomy/7145 | "Papilio" | http://ddbj.nig.ac.jp/ontologies/taxonomy/Genus |
| http://identifiers.org/taxonomy/50557 | "Insecta" | http://ddbj.nig.ac.jp/ontologies/taxonomy/Class |
| http://identifiers.org/taxonomy/76193 | "Papilio machaon" | http://ddbj.nig.ac.jp/ontologies/taxonomy/Species |

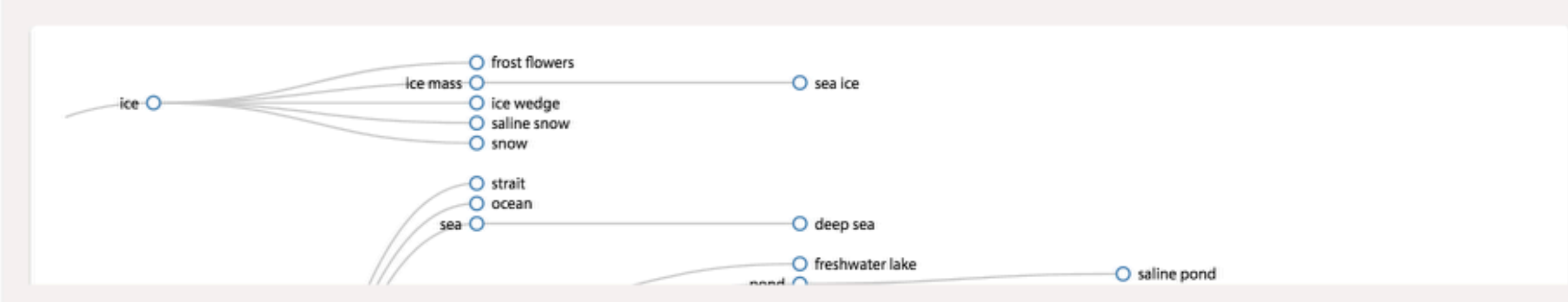
RDF guide line we propose:
<https://github.com/dbcls/rdfizing-db-guidelines/blob/master/RDF-portal-guidelines-en.md>
What type of ontology exists? → See <https://www.ebi.ac.uk/miriam/main/collections>
(identifiers.org)

TogoStanza: visualize parsed data

← → ↻ ⓘ 保護されていない通信 | togostanza.org/showcase/

Environmental ontology


Stanza Tree Environment DBCLS CC-BY



The diagram illustrates an environmental ontology. It features a central node 'ice' on the left, which branches out to 'ice mass', 'ice wedge', 'saline snow', and 'snow'. 'ice mass' further branches to 'frost flowers' and 'sea ice'. 'ice wedge' branches to 'sea ice'. 'saline snow' branches to 'sea ice'. 'snow' branches to 'sea ice'. 'sea ice' branches to 'sea', 'ocean', 'strait', 'deep sea', 'freshwater lake', and 'saline pond'. 'sea' branches to 'ocean' and 'strait'. 'ocean' branches to 'strait'. 'strait' branches to 'deep sea'. 'deep sea' branches to 'freshwater lake'. 'freshwater lake' branches to 'saline pond'.

Geographical map

Stanza Map Environment DBCLS CC-BY



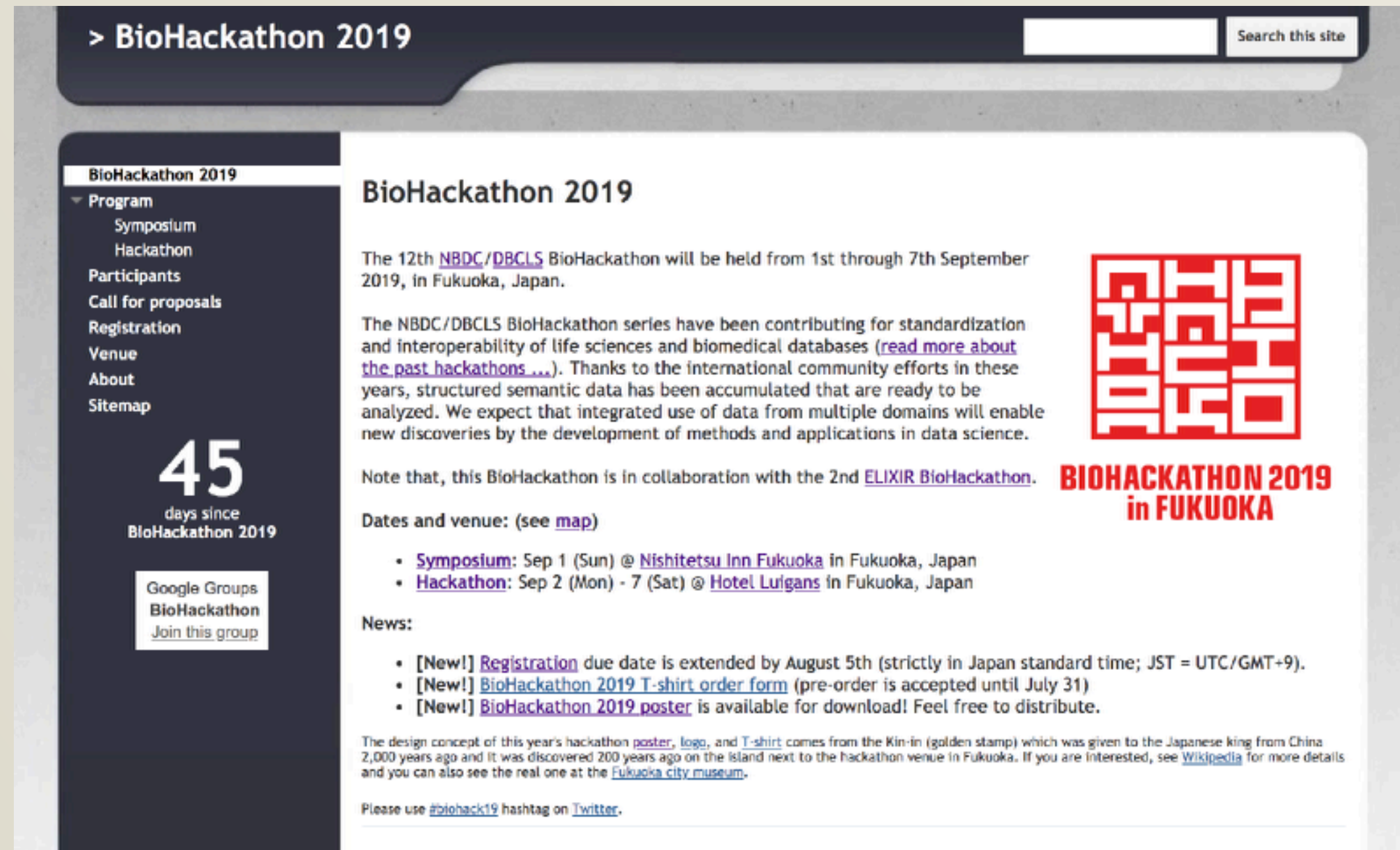
The map shows a geographical view of the Arctic region. A blue location pin is placed on the map, indicating a specific area of interest. The map includes a zoom control with '+' and '-' buttons.

Inhabitants

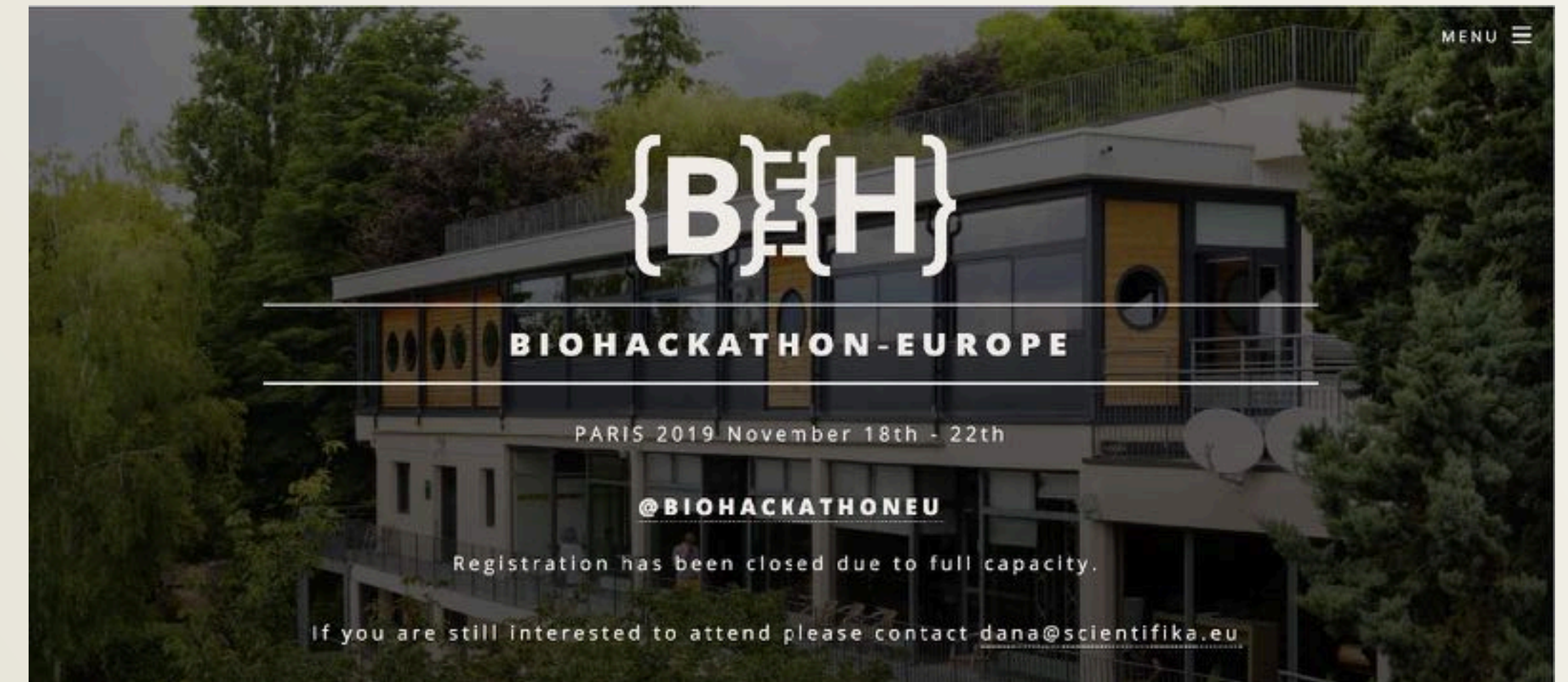
Stanza Table Environment DBCLS CC-BY

Filter (50 stanzas / 50)

Step 3: Join our BioHackathon



<http://biohackathon.org/>

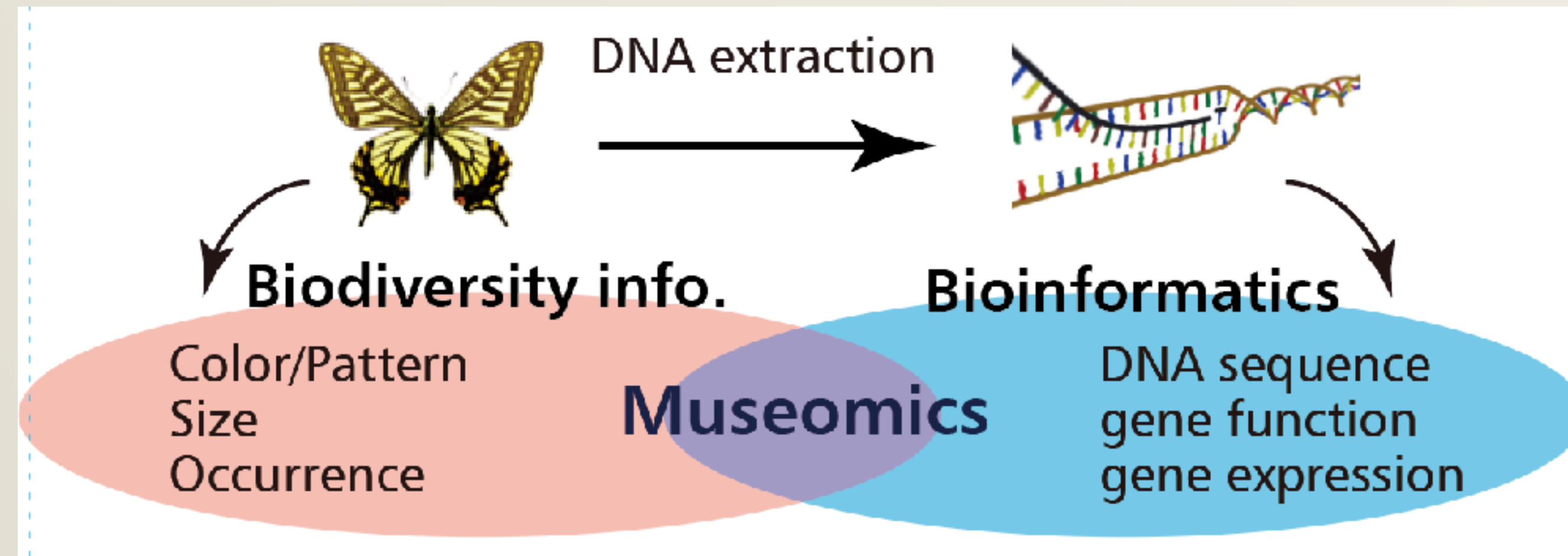


<https://www.biohackathon-europe.org/>

Annual hackathon mainly focusing on standardization and interoperability of life science data.

Go TogoTV site (<http://togotv.dbcls.jp/>) and search for "biohackathon"

Why I came here?



Museomics is an approach to obtain genetic information from preserved museum specimens by extracting DNA.

Biodiversity informatics and bioinformatics can cooperate with each other!