Unleash the Potential of your Website!

180,000 webpages from the French NHM marked up with Bioschemas/Schema.org biodiversity types

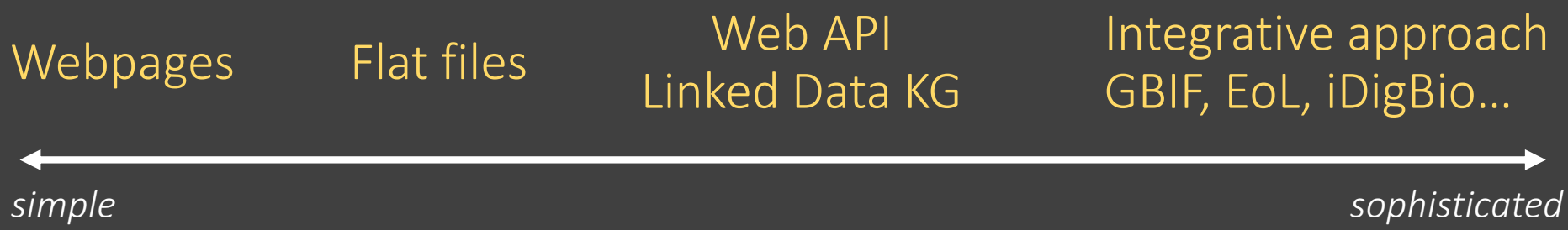
F. Michel[1], O. Gargominy[2], B. Ledentec[2], The Bioschemas Community

[1]Wimmics joint research team, Univ. Côte d'Azur, CNRS, Inria, France*
[2]UMR Patrinat, Muséum national d'Histoire naturelle, Paris, France

https://doi.org/10.3897/biss.4.59046

** Wimmics: AI in bridging social semantics and formal semantics on the Web*

How to share your biodiversity data?



Franck MICHEL - Université Côte d'Azur, CNRS, Inria, I3S, France

How to share your biodiversity data?

schema.org : semantic markup for resources on the internet

Collaborative community project founded by

Define a common vocabulary to markup resources

- Structured data makes resources understandable to search engines
- Improve ranking, discoverability
- Provide informative summarizations

Markup formats

Microformats

RDFa

Microdata

Bioschemas: extension for Life Sciences

Community initiative built on top of Schema.org

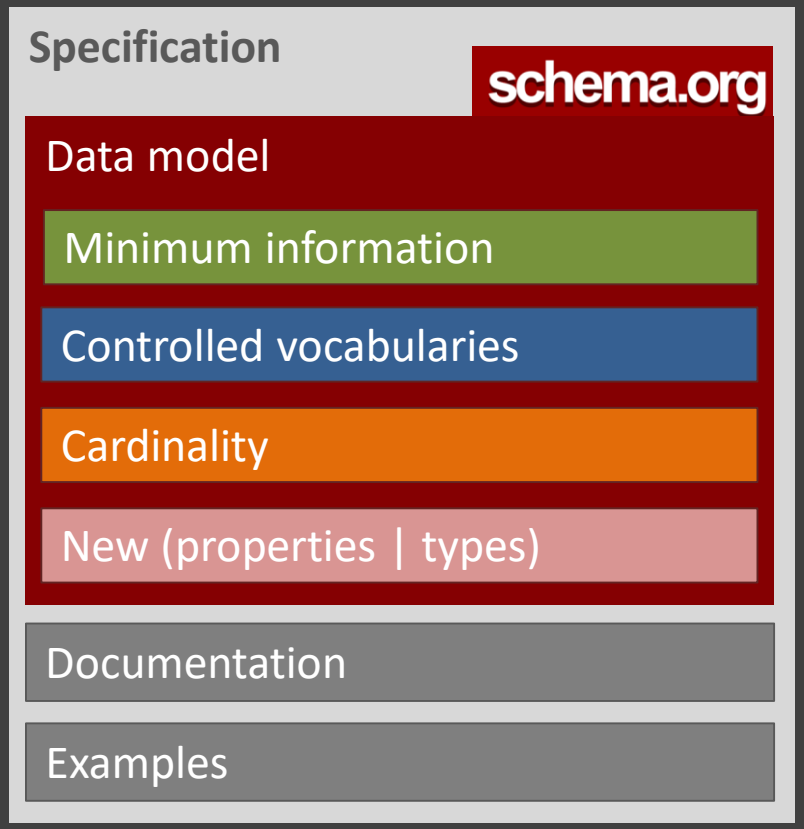
Aim

- Improve resources discoverability and interoperability in LS

Approach

- Keep it simple (no complex domain ontology)
- Provide guidelines on how to markup resources
  - Minimum/recommended/optional properties
  - Link to other vocabularies & domain ontologies

Support software



Taxon (type + profile)

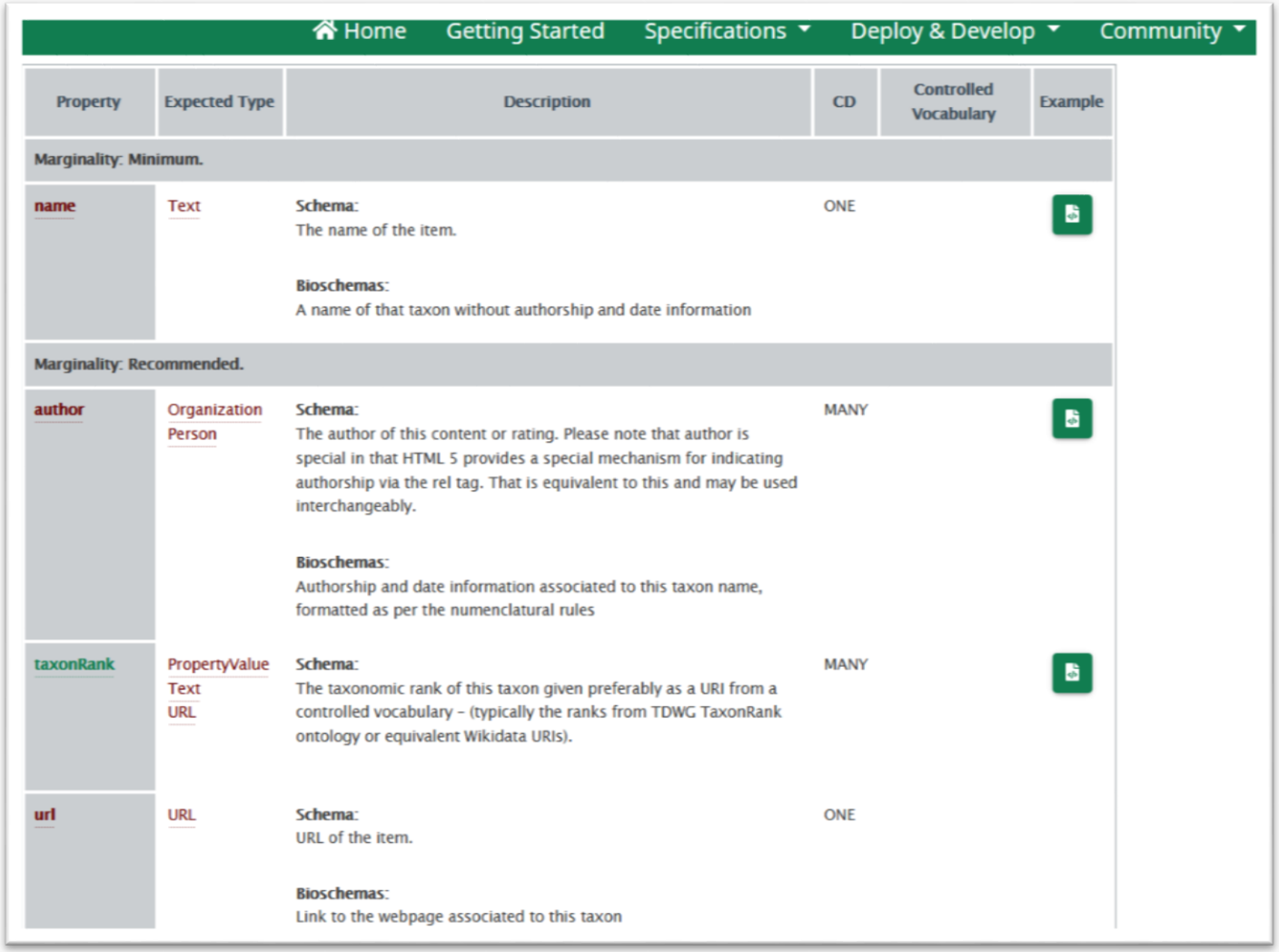
https://bioschemas.org/types/Taxon
https://bioschemas.org/profiles/Taxon

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
name	Text	**Schema**: The name of the item. **Bioschemas**: Currently valid (zoological) or accepted (botanical) name for that taxon, with authorship and date information if known.			
taxonRank	PropertyValue Text URL	**Schema**: The taxonomic rank of this taxon given preferably as a URI from controlled vocabulary (e.g. the ranks from TDWG TaxonRank ontology or equivalent Wikidata URIs)			
Marginality: Recommended.					
parentTaxon	Taxon Text URL	**Schema**: Closest parent taxon of the taxon in question. Inverse property: childTaxon **Bioschemas**: Direct, most proximate higher-rank parent taxon			
scientificName	TaxonName	**Schema**: A TaxonName representing the currently valid (zoological) or accepted (botanical) name for that taxon			

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Optional.					
additionalType	URL	**Schema**: An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax - the 'typeof' attribute - for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally. **Bioschemas**: A Taxon type from a well known vocabulary, e.g. DarwinCore http://rs.tdwg.org/dwc/terms/Taxon or http://rs.tdwg.org/ontology/voc/TaxonConcept#TaxonConcept	MANY		
alternateName	Text	**Schema**: An alias for the item. **Bioschemas**: Scientific name, with authorship and date information if known, of a synonym of the currently valid (zoological) or accepted (botanical) name.	MANY		
alternateScientificName	TaxonName	**Schema**: A TaxonName representing a scientific name, with authorship and date information if known, of a synonym of the currently valid (zoological) or accepted (botanical) name	MANY		
childTaxon	Taxon Text URL	**Schema**: Closest child taxa of the taxon in question. Inverse property: parentTaxon **Bioschemas**: Direct, most proximate lower-rank child taxa	MANY		
description	Text	**Schema**: A description of the item.			
dwc:vernacularName	Text	**Bioschemas**: A vernacular (common) name of the taxon	MANY		

TaxonName (type + profile)

https://bioschemas.org/types/TaxonName
https://bioschemas.org/profiles/TaxonName



Home Getting Started Specifications Deploy & Develop Community

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
name	Text	**Schema:** The name of the item. **Bioschemas:** A name of that taxon without authorship and date information	ONE		
Marginality: Recommended.					
author	Organization Person	**Schema:** The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably. **Bioschemas:** Authorship and date information associated to this taxon name, formatted as per the numenclatural rules	MANY		
taxonRank	PropertyValue Text URL	**Schema:** The taxonomic rank of this taxon given preferably as a URI from a controlled vocabulary – (typically the ranks from TDWG TaxonRank ontology or equivalent Wikidata URIs).	MANY		
url	URL	**Schema:** URL of the item. **Bioschemas:** Link to the webpage associated to this taxon	ONE		

Taxon and TaxonName #309:
https://github.com/BioSchemas/specifications/issues/309

Bioschemas work on biodiversity

Currently:

Taxon, TaxonName

Links to DwC terms

Future

Specimen

Links to ABCD, Open Digital Specimen?

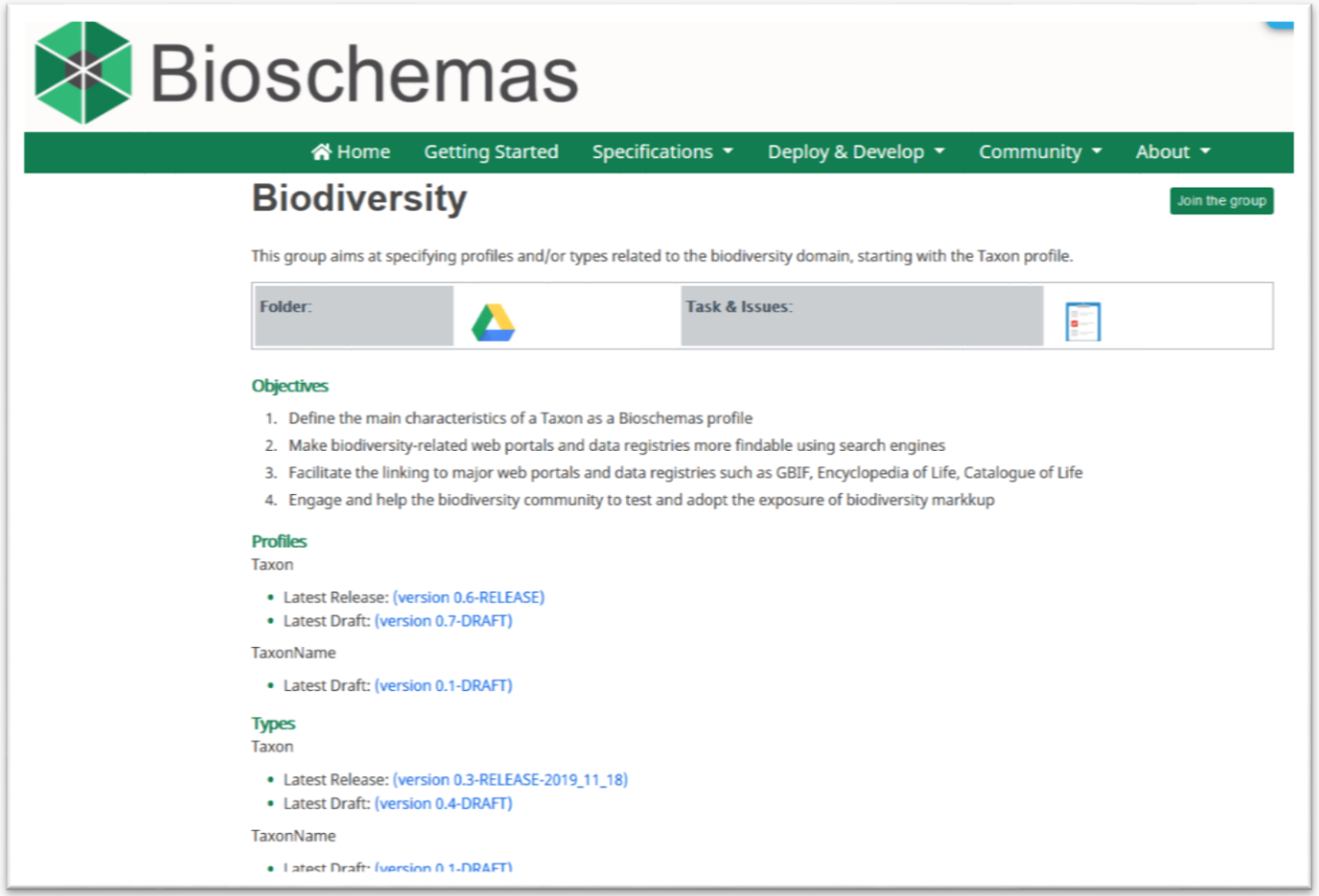
Traits

Links to traits ontologies?

Occurrence

Links to DwC occurrences?

...

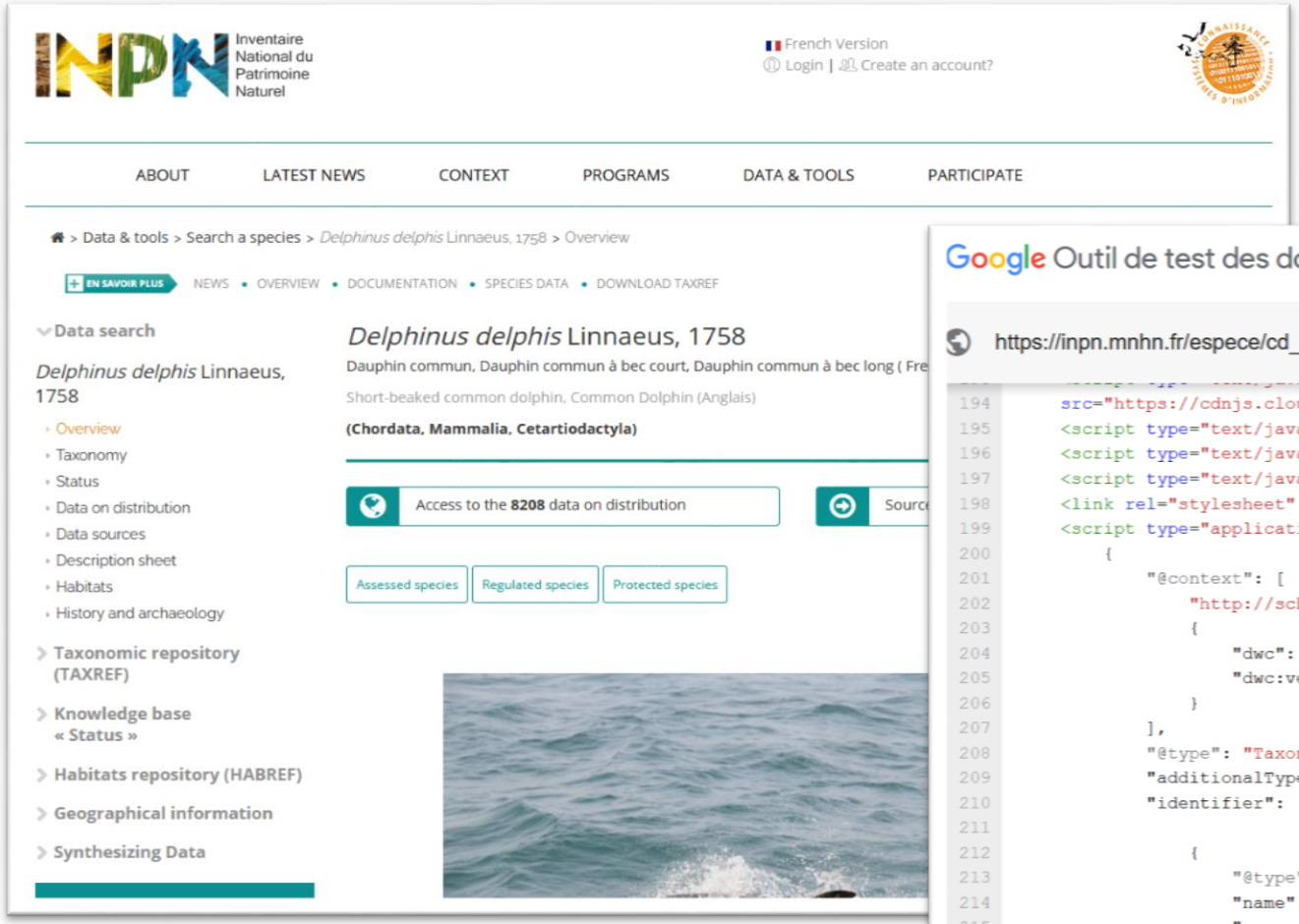


https://bioschemas.org/groups/Biodiversity/

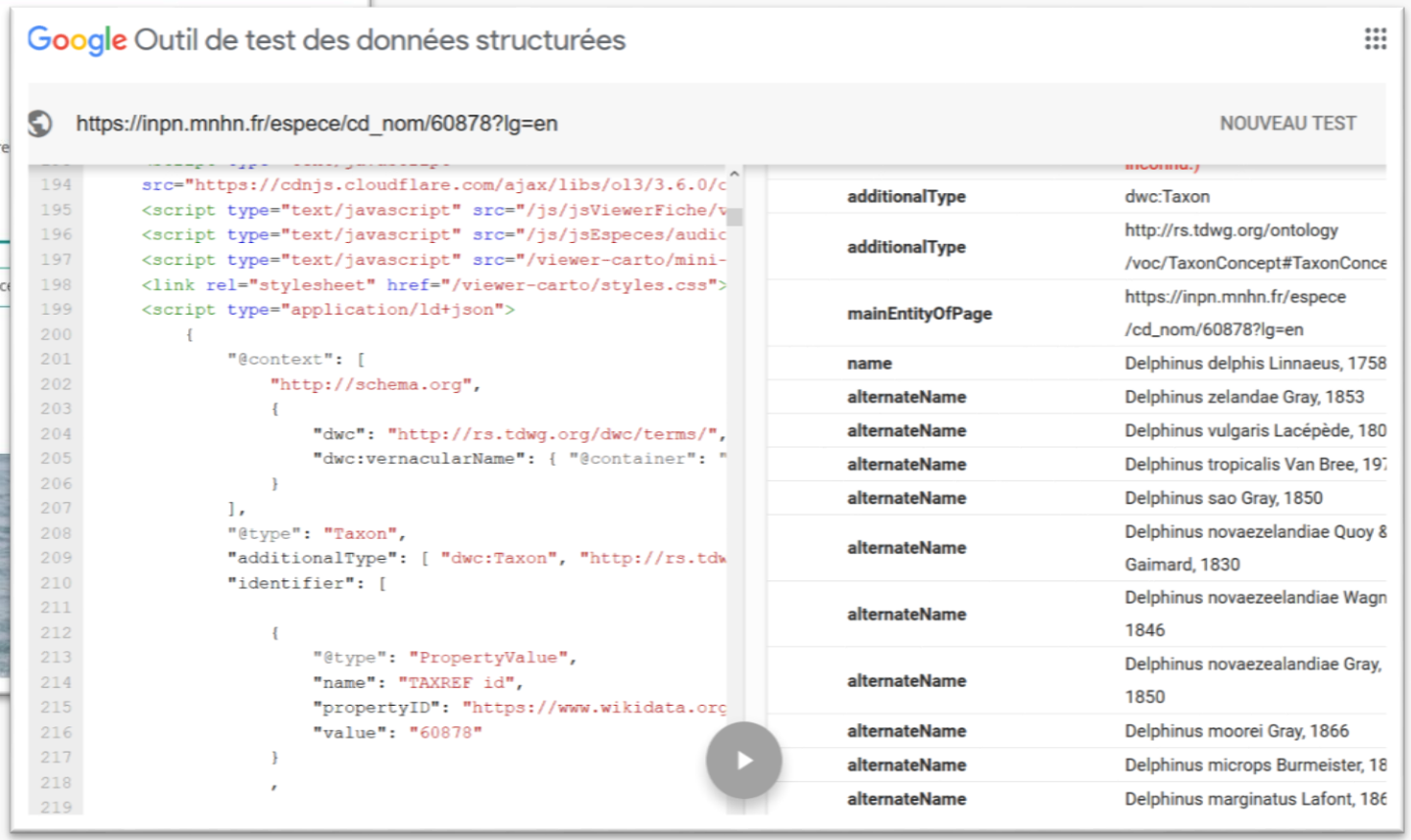
Deployment at NMNH Paris

180,000+ pages marked up with Taxon/TaxonName

Relatively inexpensive effort!

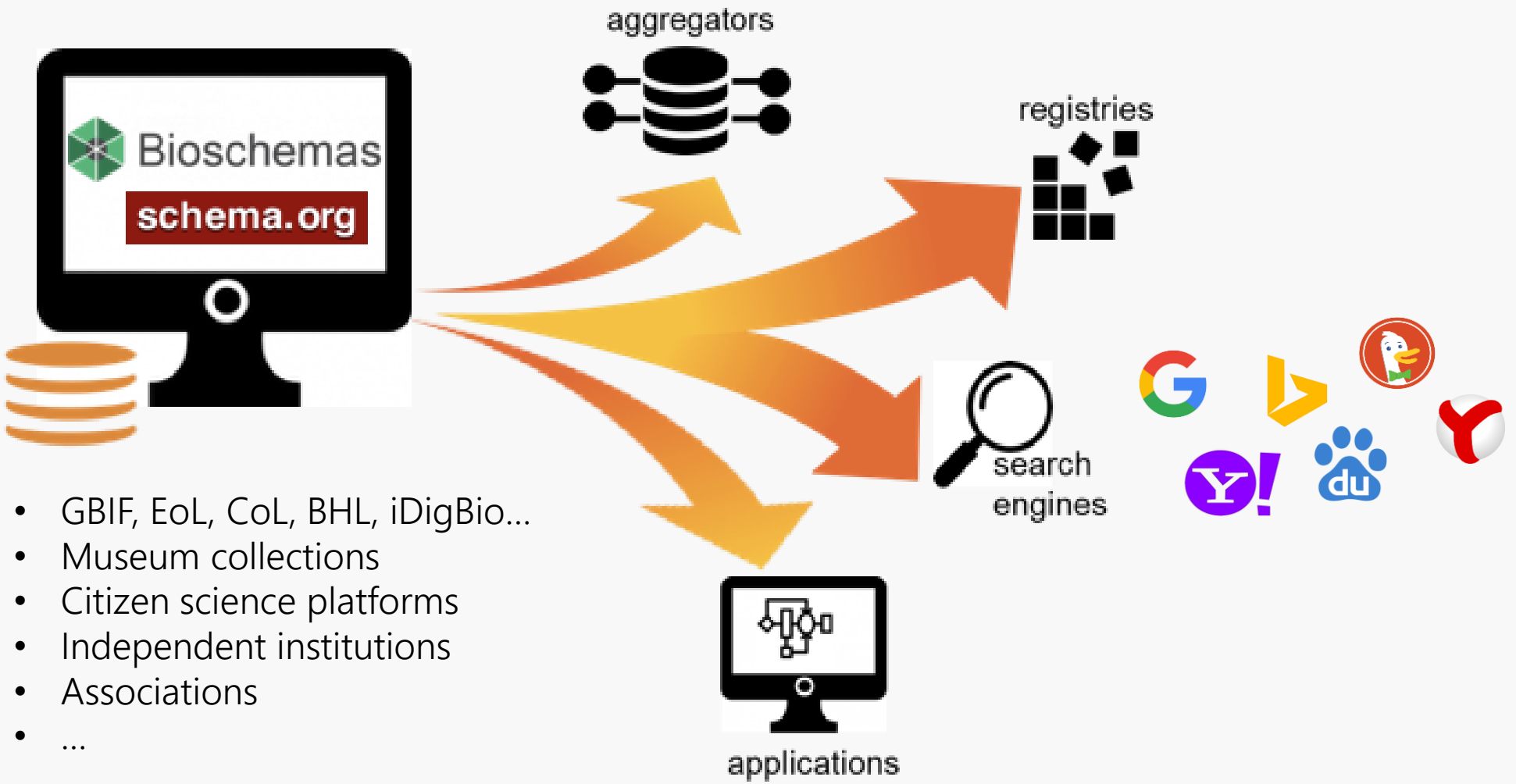


https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en



https://search.google.com/structured-data/testing-tool

Bioschemas for biodiversity… at scale



- GBIF, EoL, CoL, BHL, iDigBio…
- Museum collections
- Citizen science platforms
- Independent institutions
- Associations
- …

Take-aways

Provocative content inside!

Marking up webpages is cheap and can pay off!

- Increase discoverability, visibility
- Make “grey science” resources more visible
- …

Let’s have search engines do the job for us!

- Data integration at web scale
- Dataset search engines
- …

Thank-you

Unleash the Potential of your Website!

180,000 webpages from the French NHM marked up with Bioschemas/Schema.org biodiversity types.

F. Michel, O. Gargominy, B. Ledentec and The Bioschemas Community.

https://doi.org/10.3897/biss.4.59046

https://bioschemas.org/

Franck MICHEL - Université Côte d'Azur, CNRS, Inria, I3S, France