

Unleash the Potential of your Website!

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



F. Michel¹, O. Gargominy², B. Ledentec², The Bioschemas Community

¹Wimmics joint research team, Univ. Côte d'Azur, CNRS, Inria, France*

²UMR Patrinat, Muséum national d'Histoire naturelle, Paris, France

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

How to **share** your biodiversity data?

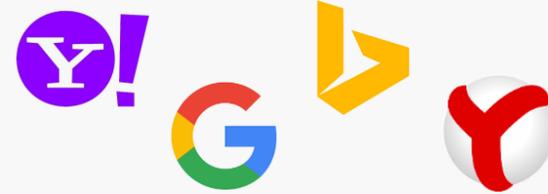


How to **share** your biodiversity data?



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



RDFa



Microformats

Microdata

Bioschemas: extension for Life Sciences

Community initiative built on top of Schema.org

Aim

Improve resources discoverability and interoperability in LS

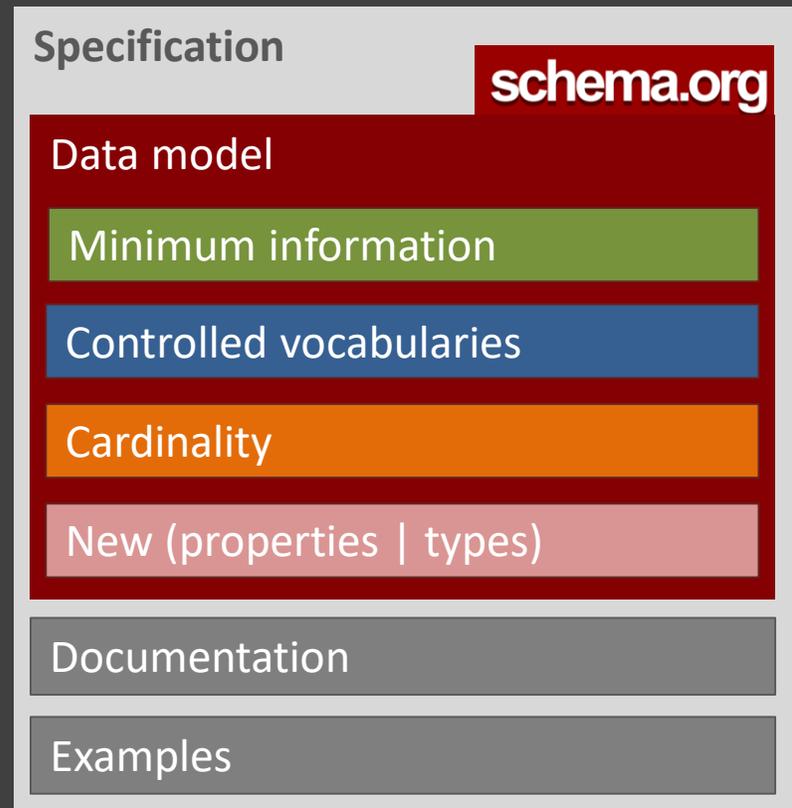
Approach

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	<p>Schema: The name of the item.</p> <p>Bioschemas: Currently valid (zoological) or accepted (botanical) name for that taxon, with authorship and date information if known.</p>			
taxonRank	PropertyValue Text URL	<p>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)</p>			
Marginality: Recommended.					
parentTaxon	Taxon Text URL	<p>Schema: Closest parent taxon of the taxon in question. Inverse property: childTaxon</p> <p>Bioschemas: Direct, most proximate higher-rank parent taxon</p>			
scientificName	TaxonName	<p>Schema: A TaxonName representing the currently valid (zoological) or accepted (botanical) name for that taxon</p>			
Marginality: Optional.					
additionalType	URL	<p>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.</p> <p>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</p>			MANY 
alternateName	Text	<p>Schema: An alias for the item.</p> <p>Bioschemas: Scientific name, with authorship and date information if known, of a synonym of the currently valid (zoological) or accepted (botanical) name.</p>			MANY 
alternateScientificName	TaxonName	<p>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</p>			MANY 
childTaxon	Taxon Text URL	<p>Schema: Closest child taxa of the taxon in question. Inverse property: parentTaxon</p> <p>Bioschemas: Direct, most proximate lower-rank child taxa</p>			MANY 
description	Text	<p>Schema: A description of the item.</p>			
dwc:vernacularName	Text	<p>Bioschemas: A vernacular (common) name of the taxon</p>			MANY 

TaxonName (type + profile)

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

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 nomenclatural 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?

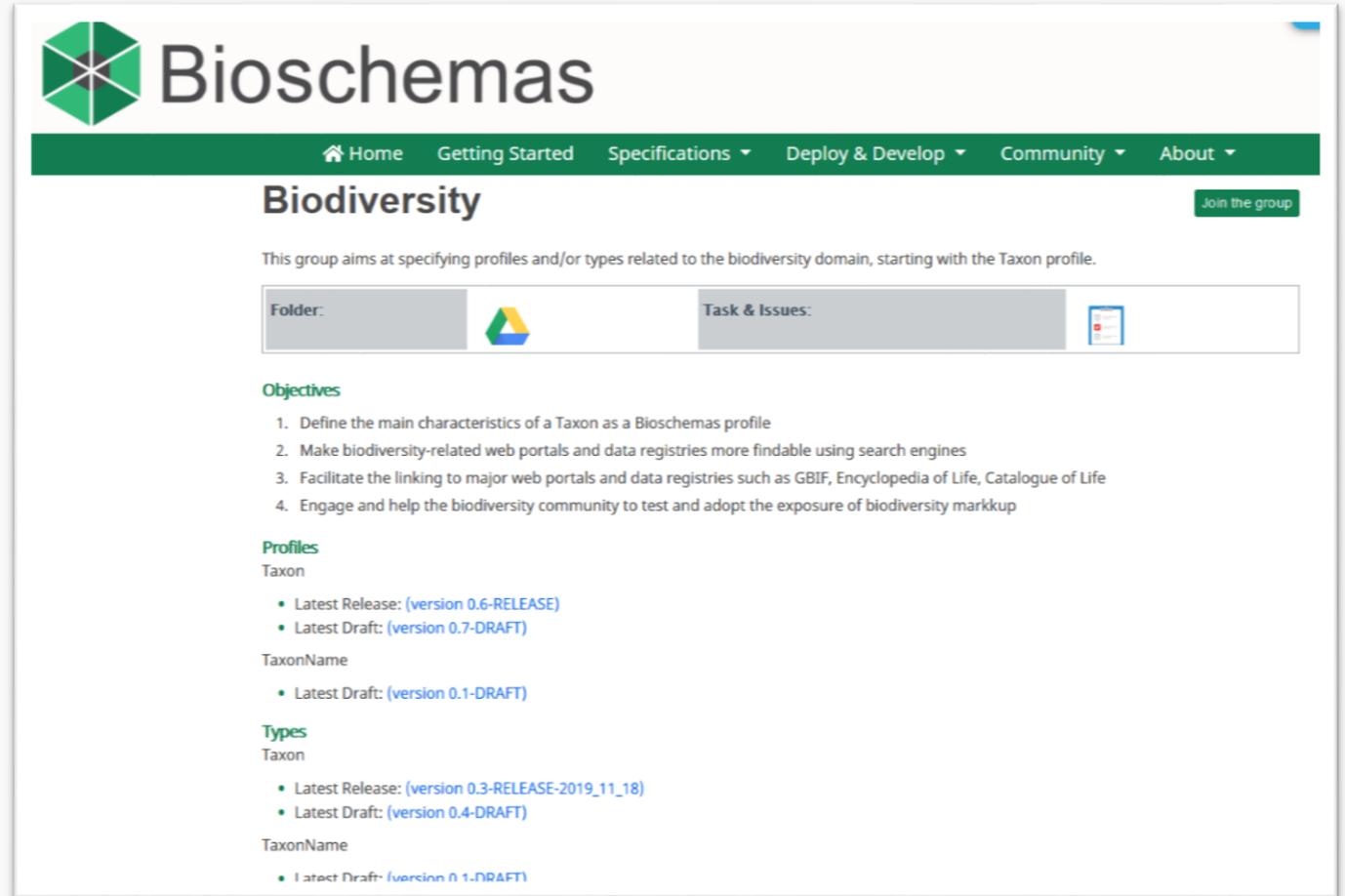
Traits

Links to traits ontologies?

Occurrence

Links to DwC occurrences?

...



The screenshot shows the Bioschemas website interface for the Biodiversity group. At the top, there is a green navigation bar with the Bioschemas logo and menu items: Home, Getting Started, Specifications, Deploy & Develop, Community, and About. Below the navigation bar, the page title is "Biodiversity" with a "Join the group" button. A description states: "This group aims at specifying profiles and/or types related to the biodiversity domain, starting with the Taxon profile." Below this, there are two sections: "Folder:" with a Google Drive icon and "Task & Issues:" with a calendar icon. The "Objectives" section lists four points: 1. Define the main characteristics of a Taxon as a Bioschemas profile; 2. Make biodiversity-related web portals and data registries more findable using search engines; 3. Facilitate the linking to major web portals and data registries such as GBIF, Encyclopedia of Life, Catalogue of Life; 4. Engage and help the biodiversity community to test and adopt the exposure of biodiversity markup. The "Profiles" section lists "Taxon" with two entries: "Latest Release: (version 0.6-RELEASE)" and "Latest Draft: (version 0.7-DRAFT)". The "Types" section lists "TaxonName" with two entries: "Latest Draft: (version 0.1-DRAFT)" and "Latest Release: (version 0.3-RELEASE-2019_11_18)".

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

Deployment at NMNH Paris

180,000+ pages marked up with Taxon/TaxonName

Relatively inexpensive effort!

The screenshot shows the INPN (Inventaire National du Patrimoine Naturel) website. The header includes the INPN logo, 'French Version', and 'Login | Create an account?'. The navigation menu has 'ABOUT', 'LATEST NEWS', 'CONTEXT', 'PROGRAMS', 'DATA & TOOLS', and 'PARTICIPATE'. The main content area is titled 'Data & tools > Search a species > Delphinus delphis Linnaeus, 1758 > Overview'. It features a search bar, a sidebar with 'Data search' options, and a main section for 'Delphinus delphis Linnaeus, 1758' with a description and a photo of a dolphin.

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

The screenshot shows the Google Structured Data Testing Tool interface. The URL is 'https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en'. The tool displays the JSON-LD data for the page, which includes a 'Taxon' type with various alternate names and a 'TaxonName' property. The JSON-LD data is as follows:

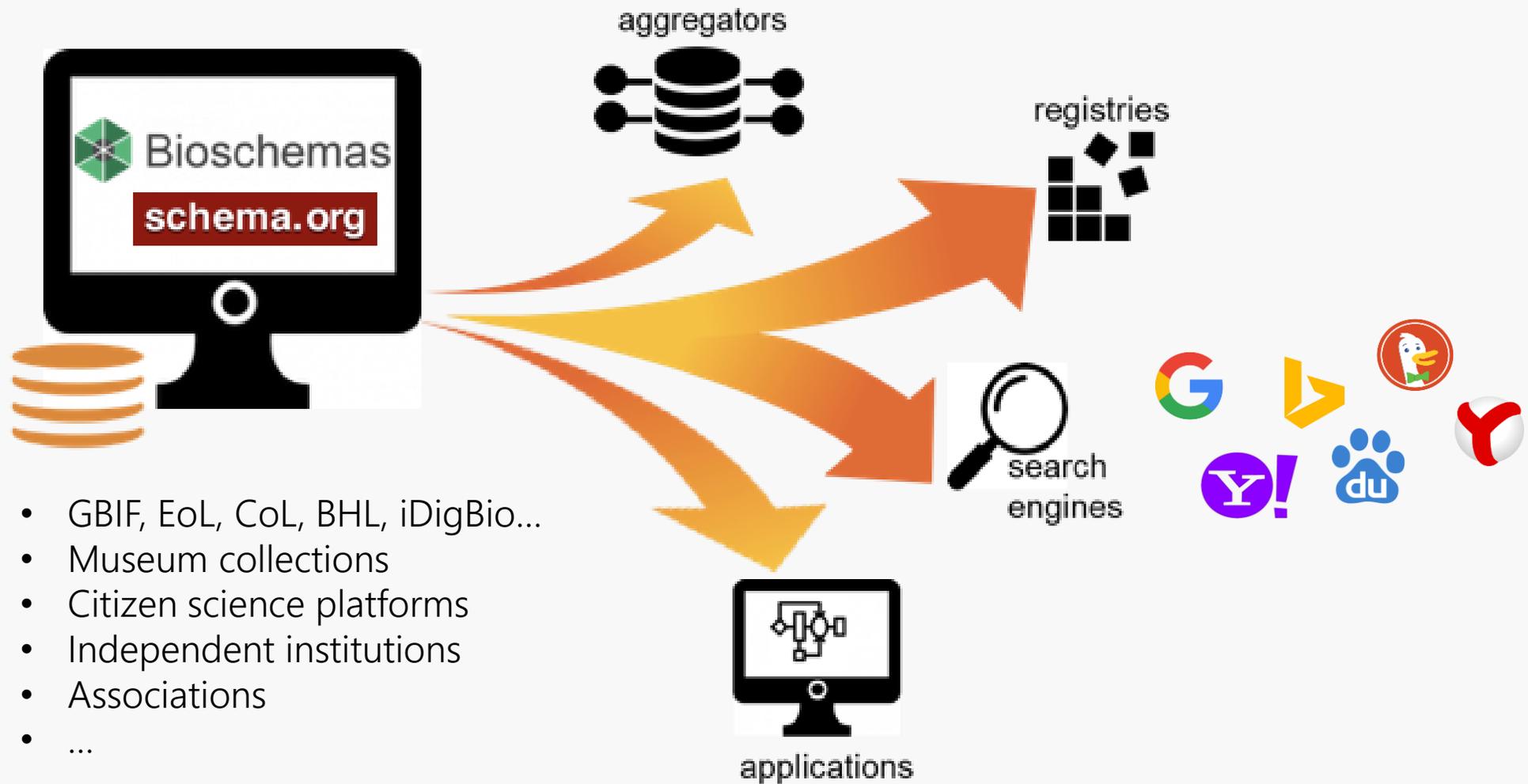
```
194   src="https://cdnjs.cloudflare.com/ajax/libs/ol3/3.6.0/c
195   <script type="text/javascript" src="/js/jsViewerFiche/v
196   <script type="text/javascript" src="/js/jsEspecies/audic
197   <script type="text/javascript" src="/viewer-carto/mini-
198   <link rel="stylesheet" href="/viewer-carto/styles.css">
199   <script type="application/ld+json">
200     {
201       "@context": [
202         "http://schema.org",
203         {
204           "dwc": "http://rs.tdwg.org/dwc/terms/",
205           "dwc:vernacularName": { "@container": "
206         }
207       ],
208       "@type": "Taxon",
209       "additionalType": [ "dwc:Taxon", "http://rs.tdw
210       "identifier": [
211         {
212           "@type": "PropertyValue",
213           "name": "TAXREF id",
214           "propertyID": "https://www.wikidata.org
215           "value": "60878"
216         }
217       ]
218     },
219
```

The tool also displays a table of structured data:

Property	Value
additionalType	dwc:Taxon
additionalType	http://rs.tdwg.org/ontology/voc/TaxonConcept#TaxonConce
mainEntityOfPage	https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en
name	Delphinus delphis Linnaeus, 1758
alternateName	Delphinus zelandae Gray, 1853
alternateName	Delphinus vulgaris Lacépède, 180
alternateName	Delphinus tropicalis Van Bree, 197
alternateName	Delphinus sao Gray, 1850
alternateName	Delphinus novaezelandiae Quoy & Gaimard, 1830
alternateName	Delphinus novaezeelandiae Wagn 1846
alternateName	Delphinus novaezeelandiae Gray, 1850
alternateName	Delphinus moorei Gray, 1866
alternateName	Delphinus microps Burmeister, 18
alternateName	Delphinus marginatus Lafont, 186

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

Bioschemas for biodiversity... at scale



Take-aways

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
- ...

Provocative content inside!

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/>