Bioschemas, structured data on your web pages

Leyla Garcia[1,3], Ricardo Arcila[1,3], Alasdair Gray[2,3], Bioschemas community[3]
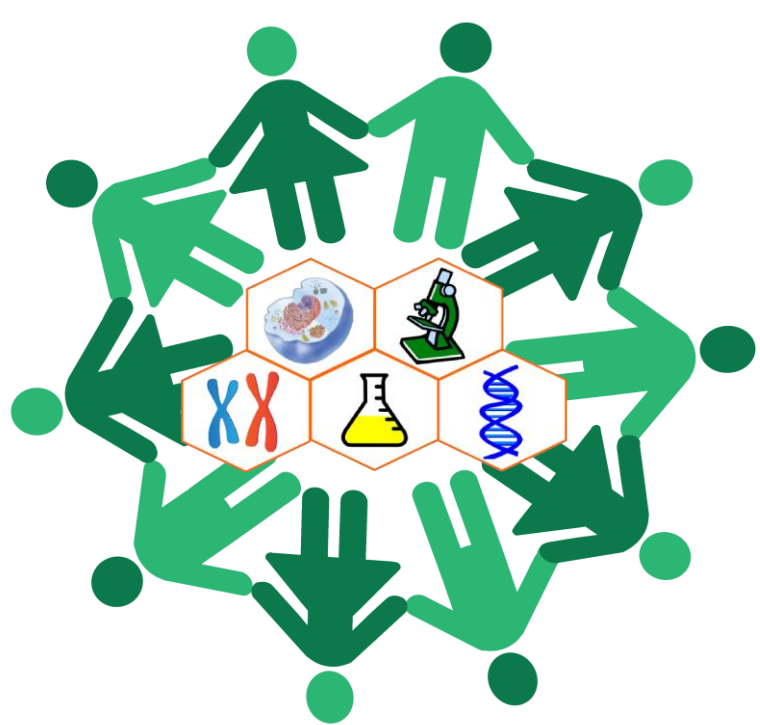
[1]EMBL-European Bioinformatics Institute, Cambridge, UK, [2] Heriot-Watt University, Edinburgh, UK, [3]http://bioschemas.org/community/
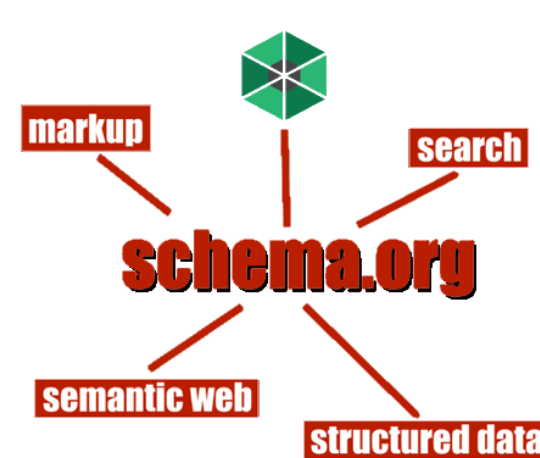
Bioschemas   **schema.org**

Bioschemas is a community initiative to extend schema.org so as to improve discoverability and interoperability of Life Sciences data exposed via websites. Bioschemas includes a variety of datatypes including datasets, samples, proteins, genes, phenotypes, diseases, laboratory protocols and phenotypes.

http://bioschemas.org/

@bioschemas

https://github.com/bioschemas/

Our approach

Propose new types to be integrated into schema.org
- BioChemEntity
- DataRecord
- LabProtocol

schema.org — markup, search, semantic web, structured data

Provide guidelines, a.k.a. profiles, about properties usage
- Minimum, recommended, optional
- Ontology reuse → properties and classes
- Cardinality → Many, one

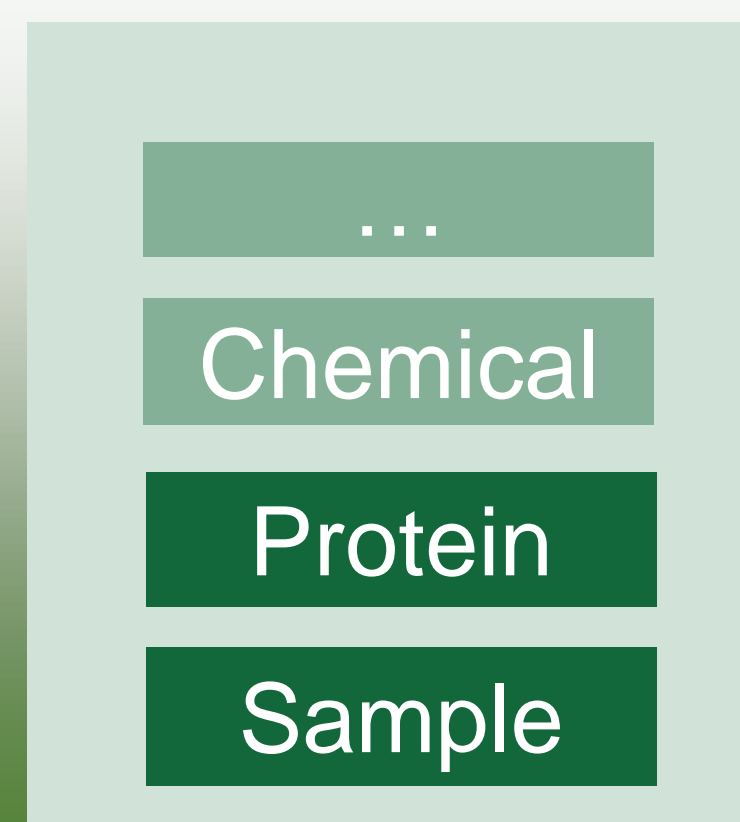## Our model at a glance: Types and profiles

Bioschemas **Profiles**

- Specific for Life Sciences
- Apply constrains to existing Schema.org types
- Minimum properties for finding and accessing data
- Best practices for selected properties
- Managed by Bioschemas

...

Chemical

Protein

Sample

DataCatalog → DataCatalog

Dataset → Dataset

... → DataRecord

Sample → BioChemEntity

LabProtocol → LabProtocol

DataCatalog → Dataset → DataRecord → BioChemEntity → LabProtocol

Schema.org **Types**

- Generic data model
- Generous list of properties to describe data types
- Managed by Schema.org

Benefits

Finding proteins, samples, phenotypes and so from resources on the web

Gathering structured information from different life science resources without dealing with multiple formats
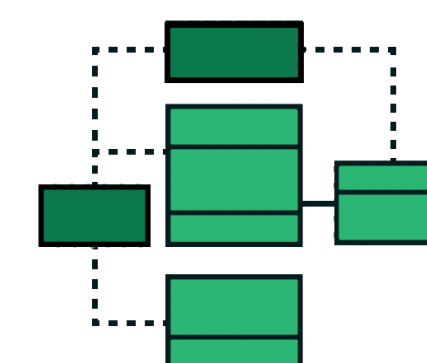
Creating quick summaries from different resources linking to the original sources

Markup preview

```
{
  "@type": "DataRecord",
  "@id": "http://www.identifiers.org/uniprot/P00519",
  "identifier": "P00519",
  "mainEntity": {
    "@type":["BioChemEntity", "Protein"],
    "identifier": "P00519",
    "alternateName": ["ABL", "JTK7"],
    "description": "Non-receptor tyrosine-protein
kinase that plays a role...",
    "name": "ABL1",
  },
  "isPartOf": {
    "@type": "Dataset",
    "@id":
"http://www.uniprot.org/news/2017/03/15/release"
  },
  "url": "http://www.uniprot.org/uniprot/P00519"
}
```
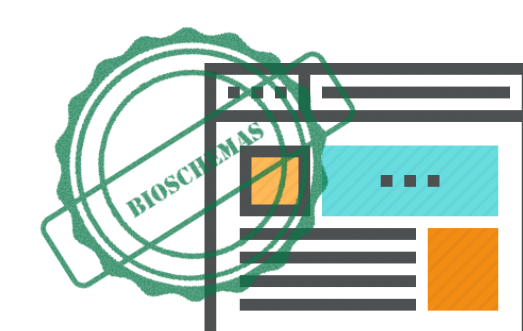
Join the community

Create new profiles, for instance phenotypes, interaction, pathway, chemical

Provide feedback via GitHub issues or create pull requests from our website

Use Bioschemas profiles to mark up elements on your Life Science web pages

Contact
public-bioschemas@w3.org
ljgarcia@ebi.ac.uk

EMBL-EBI
EBI is an outstation of the European Molecular Biology Laboratory