

Bioschemas, where schema.org meets Life Sciences

Leyla Garcia and The Bioschemas Community

Search engines make it easier for users to search and retrieve information from multiple sources. With the continuous growth of data, not only in the Life Sciences but any domain, it has become difficult for users to easily find information; search engines, therefore, need to become cleverer when presenting results to users. Schema.org is a collaborative project that aims to provide schemas so a lightweight semantic layer, in the form of mark-ups, can be added on top of web pages. This semantic mark-up makes it easier for search engines to determine whether a web page refers to a movie, a book, a product, and so on, improving the relevance of the results presented to users. Additional information can also be displayed, for instance, the number of calories or ingredients for a recipe as well as an image related to it.

Bioschemas is a community project that aims to specialize schema.org in order to include common entities used in the Life Sciences. Bioschemas reuses as much as possible, not only what already exists in schema.org but also in well-known biomedical ontologies. The strategy is simple and minimalistic: (i) from the long number of types and properties supported by schema.org, select those that are more relevant in Life Sciences for findability and summarization of data catalogs as well as their datasets and records, (ii) for common Life Science entities, define a flexible wrapper that can be later profiled according to specific needs. A profile is a pseudo-type including guidelines about marginality, cardinality and reuse of controlled vocabularies.

Bioschemas specifications comprise three new types: BioChemEntiy for biological, chemical and biochemical entities, LabProtocol for laboratory protocols and experimental records, and DataRecord for records collected in a dataset. While Bioschemas profiles cover data catalogs, datasets, laboratory protocols and bio/chem entities corresponding to proteins, protein annotations, protein structures, samples and beacons. Our community is continuously working on extending such list as well as the documentation and examples provided so more and more groups can adopt Bioschemas on their web pages.