Bioschemas: Using Schema.org to make life sciences resources more findable

Alasdair J G Gray and The Bioschemas Community[1]

The Bioschemas community (http://bioschemas.org) is developing specifications (http://bioschemas.org/specifications/) over Schema.org to enable life science resources, such as data about a specific protein, sample, or training event, to be more findable on the web. Whilst the content of well-known bioinformatic resources such as UniProt are easily findable, there is a long tail of specialist resources, such as the content of biobanks, that would benefit from embedding Schema.org markup in a standardised approach.

In this talk we describe the progress made in the Bioschemas Implementation Study (https://www.elixir-europe.org/about-us/implementation-studies/bioschemas). The study has established a thriving grassroots community that brings together interested parties from across ELIXIR and beyond (http://bioschemas.org/meetings/). The community has developed and tested several specifications for describing specific types of life sciences resources, with initial deployments being made (http://bioschemas.org/liveDeploys). This, in turn, is enabling tool developers to build proof of concept services against these specifications. For example, Identifiers.org and FAIRsharing.org can automatically harvest repository and data set descriptions for inclusion in their registries, or BioSamples can index the content of biobanks using the samples specification.

We will conclude the talk by identifying the opportunities for adoption of the Bioschemas specifications by the wider ELIXIR community, specifically core and node data resources, and the potential discoverability benefits offered by search engine indexing.

---

[1] http://bioschemas.org/people/