Life Science Organization Specification 0.1

A standard to describe and exchange organization information for Life Sciences

# Recommendation [DAY] [MONTH] 2015

This version: [github link]

Latest published version: [github link]

Previous version: [github link]

Editors: [Names]

Authors: [Names]

Contributors: [Names]

# Abstract

The Life Science Organization specification provides a way to describe bioscience organizations on the World Wide Web. It defines metadata terms that can be used in the code of web pages and applications, and builds on top of existing technologies and standards. The goal of the specification is to make it easier to discover, exchange and integrate life science organization profiles across the Internet.

# Status of this Document

This specification is under development. It is being written by a multi-institutional team from [ELIXIR](https://www.elixir-europe.org), [Pistoia Alliance](http://www.pistoiaalliance.org), [GOBLET](http://mygoblet.org), [TeSS](https://tess.elixir-uk.org), [BioSharing](https://biosharing.org) and [BBMRI](http://bbmri-eric.eu/). You can find more about the project and similar projects on the [bioschemas GitHub pages](https://github.com/BioSchemas). If you are interested in helping with this or any other of the listed projects, please visit the [Organization group](https://github.com/BioSchemas/bioschemas/wiki/Organization-Group) page.

The document will be reviewed by the [Organization Group](https://github.com/BioSchemas/bioschemas/wiki/Organization-Group) following the “[Standard Specifications Process](https://docs.google.com/document/d/1eDHBfw6frl9xAjIduLYRwcqUY3jehfzJ-xSKCc1nSsc/edit?usp=sharing)” defined by the community BioSchemas group.

The work proposed in this document builds on top of previous meetings and discussions with serveral umbrealla organisations and the feedback activelly collected by [Pistoia Alliance](http://www.pistoiaalliance.org).

# Table of Contents

[Recommendation [DAY] [MONTH] 2015](#_4h9m3cj2aibw)

[Abstract](#_iwpofulctdmg)

[Status of this Document](#_nrwnogeyt0xk)

[Table of Contents](#_1tqzfb2ey71m)

[Introduction](#_1xnqhnfcb6qv)

[Problem statement](#_uvllzvny99mb)

[Proposed solution](#_8yrtvifis3z)

[Rationale](#_l5gn97lkcv9k)

[Consensus](#_yxnpdchrh3dp)

[Adoption](#_wpa9sj4z7kju)

[Reuse](#_o93vjwbgv4kd)

[Goals](#_yam6mb62d1fa)

[Scope](#_rzhigwco4xdc)

[Data model](#_c3bnvhbso6yj)

[Data fields](#_e8iabhtclgbj)

[Controlled Vocabularies (CV)](#_hwrilg1hj6p0)

[Content Guidelines (CG)](#_aseflhc840ye)

[Cardinality](#_ugktfe6ssxha)

[Identifiers](#_d06qmi8p44tv)

[Implementation Guidelines](#_8vosxnycv8rb)

[Microdata](#_l5ii5j50d7c8)

[RDFa](#_kg4bvzc6psy9)

[JSON-LD](#_gfutfxtjn960)

[Glossary](#_s457h9uql3j5)

[Further examples](#_4ci7ip276xx5)

# Introduction

## Problem statement

Corporations, small businesses, individual consultants, academic institutes, multinationals, alliances, consortia, and informal working groups all have a part to play in the life sciences world. Each organization usually publishes information about itself on the web, and in the case of alliances and consortia that will include information about which other organizations comprise their membership, but none of this is carried out in an easily discoverable or integrated manner. This makes it hard to identify who is doing what and working with whom. Though there are technological solutions for sharing organization profiles in the life sciences at the moment there is not a consistent manner to discover, exchange or compare this information.

## Proposed solution

### Rationale

In the development of the Life Science Organization standard we have considered the following design goals.

#### *Consensus*

Many organizations and repositories providing organization profiles already exist. It is important this standard takes into account their experience and contribution.

#### *Adoption*

Many organizations already have a website or system providing information about their member organizations. They will not be willing to change their methods unless there is a clear benefit and a low barrier for adoption.

#### *Reuse*

There are existing formats and technologies suitable to represent at least some information about organizations. This specification will avoid reinvention and seek to extend existing standards.

### Goals

The Life Science Organization specification aims to support the description, discoverability, exchange and aggregation of the organization information in the life sciences. It will do this by working with the community to reach consensus on how to identify, describe and classify life science organization profiles. Components will include:

* a data model,
* a minimum information guideline,
* controlled vocabularies, and
* tools.

The specification is designed to be unintrusive minimising changes to the methods organizations currently use to publicize their profiles. It aims to facilitate adoption by extending existing standards. The definition and classification of fields in the data model use standard specifications from Schema.org, and the dissemination of information is facilitated by making use of standards like Microdata, JSON-LD and RDFa. Fields that require controlled vocabularies will specify existing ontologies where possible.

### Scope

The document is intended for:

* **software developers** who are working on projects that need to make available or aggregate organization profiles, and
* **users** who want to understand how this profiles can be described and discovered using this standard.

# Data model

The data model is based on the standards set out in Schema.org. [Schema.org](http://schema.org/) is a collaborative, community-driven project with a mission to create, maintain and promote schemas (types) for structured data on the Internet. These types (like Event, Person, Book) provide a standard for creating semantic markup in web pages and applications.

Schema.org markup covers entities, relationships between entities and actions, and can easily be extended through a well-defined extension model. Over 10 million sites already use Schema.org to code their web pages, email messages, etc. Many applications from Google, Microsoft, Pinterest, Yandex and others also use Schema.org types.

The data model proposed involves:

1. **Adopting the Schema.org Organization type, and extending it with additional properties**. Schema.org already has a way of describing organization, through its [Organization type](https://schema.org/Organization). In this document we suggest using this type to describe life science organizations, but we also suggest new properties for this type, so that organization descriptions can be more accurate and useful in life sciences. If the community agrees, these additional properties will be put forward for adoption to Schema.org.
2. **Adopting a standard way of using the Schema.org Organization type.** Many properties in the Schema.org Organization type are loosely defined, and we propose guidelines on how to use them so that they are more specific and consistent.These guidelines include concepts not supported by Schema.org, such as cardinality, controlled vocabularies and content guidelines (minimum, optional and recommended fields). For example, we suggest the use of a controlled vocabulary based on the EDAM ontology for the ‘topics’ property. These recommendations will not be part of the Schema.org Organization type, but are proposed as best practices in using that type in life science.

In the table below, the existing Schema.org properties for the Organization type are listed first, and the suggested new properties are listed afterwards (highlighted in blue). The research behind these new properties is compiled into [several documents](https://drive.google.com/folderview?id=0B1TdgUL4-iBaOXVaZ0szWlRQc2M&usp=sharing).

[Thing](https://schema.org/Thing) > [Organization](https://schema.org/Organization)

### Data fields

**Legend:**

*CN: Cardinality: one, many*

*CG: Content Guideline; M: minimum; R: recommended; O: optional*

*CV: Controlled Vocabulary*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Property** | **Expected Type** | **Description** | **CN** | **CG** | **CV** |
| **Properties inherited from** [**schema.org/Organization**](https://schema.org/Organization) | | | | | |
| address | PostalAddress | Physical address of the item. | One |  |  |
| aggregateRating | AggregateRating | *(Not anticipated to be used in this context)* The overall rating, based on a collection of reviews or ratings, of the item. | One |  |  |
| award | Text | An award won by or for this item. Supersedes awards. | Many |  |  |
| brand | Brand or Organization | The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person. | Many |  |  |
| contactPoint | ContactPoint | A contact point for a person or organization. Supersedes contactPoints. | Many | R |  |
| department | Organization | A relationship between an organization and a department of that organization, also described as an organization (allowing different urls, logos, opening hours). For example: a store with a pharmacy, or a bakery with a cafe. | Many | O |  |
| dissolutionDate | Date | The date that this organization was dissolved. | One | O |  |
| duns | Text | The Dun & Bradstreet DUNS number for identifying an organization or business person. | One |  |  |
| email | Text | Email address. | One |  |  |
| employee | Person | *(Not anticipated to be used in this context)* Someone working for this organization. Supersedes employees. | Many |  |  |
| event | Event | Upcoming or past event associated with this place, organization, or action. Supersedes events. | Many |  |  |
| faxNumber | Text | The fax number. | One |  |  |
| founder | Person | A person who founded this organization. Supersedes founders. | Many |  |  |
| foundingDate | Date | The date that this organization was founded. | One |  |  |
| foundingLocation | Place | The place where the Organization was founded. | One |  |  |
| globalLocationNumber | Text | The Global Location Number (GLN, sometimes also referred to as International Location Number or ILN) of the respective organization, person, or place. The GLN is a 13-digit number used to identify parties and physical locations. | One |  |  |
| hasPOS | Place | Points-of-Sales operated by the organization or person. | Many |  |  |
| isicV4 | Text | The International Standard of Industrial Classification of All Economic Activities (ISIC), Revision 4 code for a particular organization, business person, or place. | One |  |  |
| legalName | Text | The official name of the organization, e.g. the registered company name. | One | M |  |
| location | Place or PostalAddress | The location of the event, organization or action. | One | R |  |
| logo | URL or ImageObject | An associated logo. | One | R |  |
| makesOffer | Offer | A pointer to products or services offered by the organization or person. | Many |  |  |
| member | Person or Organization | A member of an Organization or a ProgramMembership. Organizations can be members of organizations; ProgramMembership is typically for individuals. Supersedes musicGroupMember, members. Inverse property: memberOf. | Many | R |  |
| memberOf | ProgramMembership or Organization | An Organization (or ProgramMembership) to which this Person or Organization belongs. Inverse property: member. | Many | R |  |
| naics | Text | The North American Industry Classification System (NAICS) code for a particular organization or business person. | One |  |  |
| numberOfEmployees | QuantitativeValue | The number of employees in an organization e.g. business. | One |  |  |
| owns | OwnershipInfo or Products | Products owned by the organization or person. | Many | O |  |
| parentOrganization | Organization | The larger organization that this organization is a branch of, if any. Supersedes branchOf. Inverse property: subOrganization. | One | O |  |
| review | Review | *(Not anticipated to be used in this context)* A review of the item. Supersedes reviews. | Many |  |  |
| seeks | Demand | A pointer to products or services sought by the organization or person (demand). | Many |  |  |
| subOrganization | Organization | A relationship between two organizations where the first includes the second, e.g., as a subsidiary. See also: the more specific 'department' property. Inverse property: parentOrganization. | Many | O |  |
| taxID | Text | The Tax / Fiscal ID of the organization or person, e.g. the TIN in the US or the CIF/NIF in Spain. | One |  |  |
| telephone | Text | The telephone number. | One |  |  |
| vatID | Text | The Value-added Tax ID of the organization or person. | One |  |  |
| **Proposed new properties for the Organization type** | | | | | |
| type | CV | The type of the organization selected from a controlled vocabulary: single-company corporation, multinational, not-for-profit, alliance, consortium, institute, department, working party, project, etc. | Many | M | X |
| dateFounded | [Date](https://schema.org/Date) | The date the organization was originally founded. If created through a merger, this should be the foundation date of the purchasing organization. | One |  |  |
| founderMember | Person or Organization | For an alliance or collaboration or project, the founding members. | Many |  |  |
| status | CV and/or Text | To record if the organization is still active, and if it were a project or alliance, what the outcome was, or if it has closed down or merged, why and when. | One | R | X |
| budget | PriceSpecification | The annual (or total - for a project) budget for the organization. | One |  |  |
| fundingModel | CV and/or Text | A description of how the organization is funded. | One | R | X |
| membershipCategory | Offer | Defines a membership category and associated fees (for projects and alliances etc.) | Many | R |  |
| attachment | [URL](https://schema.org/URL) | Any files or related websites which give more information about this organization. e.g flyers, third party sites, ... | Many |  |  |
| socialMedia | [URL](https://schema.org/URL) | Link to social media websites like twitter or facebook. | Many |  |  |
| lastUpdate | [Date](https://schema.org/Date) | Date when the organization information was last modified | One |  |  |
| topic | CV and/or Text | Scientific topics that describe the organization and its activities, using a scientific ontology term from EDAM topics. | Many | M | X |
| keywords | CV and/or [Text](https://schema.org/Text) | Keywords to describe the organization. Use term which are not available in EDAM “Scientific Topic”. Use text keywords or ontology terms from other ontologies which could complement EDAM topics. | Many | R |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Properties inherited from** [**schema.org/Thing**](https://schema.org/Thing) | | | | | |
| [additionalType](https://schema.org/additionalType) | [URL](https://schema.org/URL) | 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. | Many |  |  |
| [alternateName](https://schema.org/alternateName) | [Text](https://schema.org/Text) | An alias for the item. | One | R |  |
| [description](https://schema.org/description) | [Text](https://schema.org/Text) | A short description of the item. | One | M |  |
| image | ImageObject or URL | *(At least one of these should be the organisation’s logo, if it has one.)* An image of the item. This can be a URL or a fully described ImageObject.Inverse property: mainEntity. | Many |  |  |
| mainEntityOfPage | CreativeWork or URL | Indicates a page (or other CreativeWork) for which this thing is the main entity being described. See background notes for details. | One |  |  |
| [name](https://schema.org/name) | [Text](https://schema.org/Text) | The name of the item. | One | M |  |
| [potentialAction](https://schema.org/potentialAction) | [Action](https://schema.org/Action) | Indicates a potential Action, which describes an idealized action in which this thing would play an 'object' role. | Many |  |  |
| [sameAs](https://schema.org/sameAs) | [URL](https://schema.org/URL) | URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Freebase page, or official website. | One | M |  |
| [url](https://schema.org/url) | [URL](https://schema.org/URL) | URL of the item. | Many | R |  |

Here is an example snippet showing the attributes you can use on an Organization. For complete examples see the “Further examples” section at the end of this document.

|  |
| --- |
| **Example 1. Attributes/properties of an Organization** |
| <div itemscope itemtype="http://schema.org/Organization">  <span itemprop="name">The European Bioinformatics Institute</span> (<span itemprop="legalName">EMBL-EBI</span>)  <div itemprop="description">  Centre for research and services in bioinformatics and part of European Molecular Biology Laboratory (EMBL).  </div>   * Contact Details:   <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">  Main address:  <span itemprop="streetAddress">Wellcome Genome Campus</span>   * <span itemprop="postalCode">CB10 1SD</span> * <span itemprop="addressLocality">Hinxton, Cambridge</span>   </div>  Tel:<span itemprop="telephone">+44 (0)1223 494 444</span>,  Fax:<span itemprop="faxNumber">+44 (0)1223 494 468</span>  </div>  <a itemprop="url" href="http://www.ebi.ac.uk">http://www.ebi.ac.uk</a>  <link itemprop="sameAs" href="https://en.wikipedia.org/wiki/European\_Bioinformatics\_Institute"/>  <div>Type:  <span itemprop="type">institute</span>,  <span itemprop="type">not-for-profit</span>  </div>  <div>Topics:  <span itemprop="topic">Bioinformatics</span>,  <span itemprop="topic">Computational biology</span>,  <span itemprop="topic">Biology</span>  </div>  </div> |

### Controlled Vocabularies (CV)

Some data fields suggest the use of controlled vocabularies or enumerations. We will rely on existing vocabularies and ontologies wherever possible but define new collections of terms for very specific purposes as required.

This section contains a list of fields that require a controlled vocabulary, enumeration or an ontology term, and specifies what is acceptable for each. The fields involved are:

* **Type**  
  *Could be* one of:

|  |  |
| --- | --- |
| **Value** | **Description** |
| single-company corporation |  |
| multinational |  |
| not-for-profit |  |
| alliance |  |
| consortium |  |
| institute |  |
| department |  |
| working party |  |
| project |  |
| ... |  |

* **Status**  
  *Could be* one of:

|  |  |
| --- | --- |
| **Value** | **Description** |
| active |  |
| dormant |  |
| merged |  |
| closed |  |
| ... |  |

* **Funding Model**  
  ...
* **Topic**  
  *Must be* one of the [EDAM Topic](http://edamontology.org/topic_0003) class values.

### Content Guidelines (CG)

To make it as easy as possible to implement a basic Life Science Organization model, we suggest a very small set of minimum (M) fields to include. For optimal discovery and integration we suggest some additional recommended (R) fields. All other fields are optional (O), but if included will enhance the user experience.

Fields that *must be* present (M) in order to comply with the specification are:

* type
* topic
* legalName
* description
* name
* sameAs

### Cardinality

The Schema.org specification permits any field to be included any number of times. Whether this is desirable depends on the context and intended use of the data. This specification includes suggestions as to the cardinality of selected fields, as indicated in the data model table above.

The table notates cardinalities in the following way:

|  |  |
| --- | --- |
| **Notation** | **Definition** |
| One | There may only be a maximum of one instance of this property type. For example, an organization may only have a maximum of one legal name. |
| Many | There can be multiple instances of this property type. For example, there may be more than one contact points for an organization. |

|  |
| --- |
| **Example 2. Cardinality in Life Science Organization properties as microdata within HTML** |
| ... |

*An example of ...*

### Identifiers

TBD

## Implementation Guidelines

Schema.org [suggests](http://schema.org/docs/gs.html) implementing metadata, including the Life Science Organization specification, using Microdata, RDFa, or JSON-LD. Depending on the context, any of these can be used for embedding compliant organization profiles in organisations web pages or other online resources and services.

### Microdata

Microdata can be used for embedding properties from the specification directly into existing web pages and HTML tags to enrich organization profile descriptions. This microdata can be extracted and further processed by search engines and other applications, but does not affect the ‘look and feel’ of the web page it is embedded in. Using microdata is the easiest method of implementing the specification, as it requires minimal intervention. Example below depicts the use of microdata within HTML tags.

|  |
| --- |
| **Example 3. Embedding Life Science Organization properties as microdata within HTML** |
| ... |

For more information, please refer to the [Microdata Guide on Schema.org](https://schema.org/docs/gs.html).

### RDFa

[RDFa](https://en.wikipedia.org/wiki/RDFa) (or [Resource Description Framework](https://en.wikipedia.org/wiki/Resource_Description_Framework) in Attributes[[1]](https://en.wikipedia.org/wiki/RDFa#cite_note-n-1)) is a [W3C](https://en.wikipedia.org/wiki/W3C) Recommendation that adds a set of attribute-level extensions to [HTML](https://en.wikipedia.org/wiki/HTML), [XHTML](https://en.wikipedia.org/wiki/XHTML) and various XML-based document types for embedding rich [metadata](https://en.wikipedia.org/wiki/Metadata) within web documents. Example below explains the use of RDFa within HTML tags.

|  |
| --- |
| **Example 4. Embedding Life Science Organization properties as RDFa within HTML** |
| ... |

For more information, please refer to the [RDFa wiki](http://rdfa.info/).

### JSON-LD

[JSON-LD](https://en.wikipedia.org/wiki/JSON-LD) (JavaScript Object Notation for Linked Data), is a method of transporting [Linked Data](https://en.wikipedia.org/wiki/Linked_Data) using [JSON](https://en.wikipedia.org/wiki/JSON). Example below represents a Life Science Organization described in JSON-LD format.

|  |
| --- |
| **Example 5. Representing Life Science Organization in JSON-LD format** |
| ... |

For more information, please refer to the [JSON-LD specification](http://www.w3.org/TR/json-ld/).

# Glossary

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Ontology/Controlled vocabulary | For the purposes of this document the terms ontology and controlled vocabulary are interchangeable. [Wikipedia](https://en.wikipedia.org/wiki/Ontology_%28information_science%29) defines ontologies as:  *“In* [*computer science*](https://en.wikipedia.org/wiki/Computer_science) *and* [*information science*](https://en.wikipedia.org/wiki/Information_science)*, an ontology is a formal naming and definition of the types, properties, and interrelationships of the* [*entities*](https://en.wikipedia.org/wiki/Entities) *that really or fundamentally exist for a particular* [*domain of discourse*](https://en.wikipedia.org/wiki/Domain_of_discourse)*. It is thus a practical application of philosophical* [*ontology*](https://en.wikipedia.org/wiki/Ontology)*, with a* [*taxonomy*](https://en.wikipedia.org/wiki/Taxonomy_%28general%29)*.”* |
| EDAM ontology | [EDAM ontology](http://edamontology.org/page) is one of the ontologies available in the life sciences domain, for classifying and describing bioinformatics operations, types of data, formats, and scientific topics. |
| EDAM ontology topic | EDAM ontology topics describe general bioinformatics subjects or categories, such as a field of study, data, processing, analysis or technology - starting from very general terms such as “biology” and “bioinformatics” to more specific ones such as "sequence analysis", "alignment", "sequencing", "microarrays", etc. |

# 

# Further examples

high-level examples…

Description of a pharma company with one HQ (e.g. GSK)

* would be just one LifeScienceOrganization record

Description of an academic institute (e.g. EBI)

* ditto

Description of an academic or industry consortium with N members

* would be N individual LifeScienceOrganization records with the “memberOf” field completed with a reference to the group’s record
* plus a LifeScienceOrganization record to represent the whole group, with N “member” fields referencing N members
  + should include at least one “membershipCategory” record detailing how to become a member
  + nb. while we can describe categories, there is no way of assigning a category for a membership relation in this data model and so we cannot indicate which category a member is, e.g. Pistoia Core or Participating

Description of a project within a group consortium

* a project is just another LifeScienceOrganization that has “parentOrganization” set to the consortium record identifier, and the consortium has a “subOrganization” field added which includes the project record identifiter
* individual companies can be members of projects as well as members of consortia (e.g. the Pistoia structure) and so the “member” and “memberOf” fields should be set on project and company records to reflect that