Tools Specification 0.1

A specification for describing life science software tools

# Abstract

The Life Science Tools specification provides a way to describe bioscience tools and software on the World Wide Web. It defines a set of metadata and vocabularies, built on top of existing technologies and standards, that can be used to represent such tools in Web pages and applications. The goal of the specification is to make it easier to discover, exchange and integrate information about life science tools across the Internet.

## Tools type definition

### 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** |
| **Existing properties in** [**schema.org/SoftwareApplication**](https://schema.org/SoftwareApplication) | | | | | |
| applicationCategory | Text | Type of software e.g. a database, tool, service. | Many | M | x |
| keywords | Text | Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas. | Many | M | x |
| featureList | Text | Features or modules provided by this application (and possibly required by other applications). | Many | M | x |
| softwareVersion | Text | Version of the software instance. | One | M |  |
| You can also use other properties from [SoftwareApplication](https://schema.org/SoftwareApplication). | | | | | |
| **Proposed new properties for the SoftwareApplication type** | | | | | |
| inputDataType | Text | Type(s) of data: primary inputs (if any), e.g. "Protein sequences" | Many | R | x |
| outputDataType | Text | Type(s) of data: primary inputs (if any), e.g. "Protein sequences" | Many | R | x |
| **Properties inherited from** [**schema.org/CreativeWork**](http://schema.org/CreativeWork) | | | | | |
| Publisher | Person or Organization | Use for contacts or ‘credit’. | Many | R |  |
| Publication | Text | Publication about this software | Many | R |  |
| license | text | The applicable software license | Many | R | ? |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Properties inherited from** [**schema.org/Thing**](https://schema.org/Thing) | | | | | |
| [description](https://schema.org/description) | [Text](https://schema.org/Text) | A short description of the item. | One | M |  |
| [name](https://schema.org/name) | [Text](https://schema.org/Text) | The name of the item. | One | M |  |
| [url](https://schema.org/url) | [URL](https://schema.org/URL) | URL of the item.This property can be used on a page listing many events, to indicate each individual event’s page. | One | M |  |

### 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:

* **applicationCategory**

Must be one of the [valid values](https://github.com/bio-tools/biotoolsschema#tool-types-v20beta03) for biotoolsSchema 2.0

* **Input data type**

Must be one of the [EDAM Data](http://edamontology.org/data_0006) concept labels or one of its synonyms.

* **Output data type**

Must be one of the [EDAM Data](http://edamontology.org/data_0006) concept labels or one of its synonyms.

* **Keywords**

*Must be* one of the [EDAM Topic](http://edamontology.org/topic_0003) concept labels or one of its synonyms.

* **featureList**

*Must be* one of the [EDAM Operation](http://edamontology.org/operation_0004) concept labels or one of its synonyms.

### Content Guidelines (CG)

To make it as easy as possible to implement a basic Tools 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:

* name
* description
* url
* applicationCategory
* keywords
* featureList
* softwareVersion

Fields that are *recommended* (R) in order to comply with the specification are:

* Input data type
* Output data type
* Publication
* accountablePerson
* license

Fields that are optional (O) in order to comply with the specification are:

### 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 event may only have a maximum of one start date. |
| Many | There can be multiple instances of this property type. For example, there may be more than one sponsor of an event. |

|  |
| --- |
| **Example 2. Cardinality in Tools properties as microdata within HTML** |
| <div itemscope itemtype="http://schema.org/SoftwareApplication">  …  <div><meta itemprop="name">ExpaRNA</div>  <div>Type:  <span itemprop="applicationCategory">Command-line tool</span>,  <span itemprop="applicationCategory">Script</span>  </div>  ...  </div> |

*An example of a property type with multiple cardinality (applicationCategory) and single cardinality (name).*

## Implementation Guidelines

Schema.org [suggests](http://schema.org/docs/gs.html) implementing metadata using JSON-LD, RDFa or Microdata. JSON-LD is the recommended format by Google but of these formats can be used for embedding information about tools in a web page or other online resource.

### 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 an Event described in JSON-LD format.

|  |
| --- |
| **Example 5. Representing a tool in JSON-LD format** |
| <script type="application/ld+json">  {  "@context": "http://schema.org/",  "@type": "SoftwareApplication",  "name": "Pladipus",  "description": "The [CompOmics](http://www.compomics.com/) Pladipus system is an affordable cluster computing framework allowing for distributed processing, management and analyses for bio-informatics pipelines.",  "applicationCategory": "Command line",  "applicationCategory": "Desktop GUI",  "keywords": "Proteomics",  "softwareVersion": "0.4.1",  ...  }  </script> |

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

### 

### Microdata

Microdata can be used for embedding properties from the specification directly into existing web pages and HTML tags to enrich tools descriptions.

|  |
| --- |
| **Example 3. Embedding Tools properties as microdata within HTML** |
| <div itemscope itemtype="http://schema.org/SoftwareApplication">  <div itemprop="name">Pladipus</div>  <div itemprop="description">The [CompOmics](http://www.compomics.com/) Pladipus system is an affordable cluster computing framework allowing for distributed processing, management and analyses for bio-informatics pipelines.</div>  <div>Application categories:  <span itemprop="applicationCategory">Command line</span>, <span itemprop="applicationCategory">Desktop GUI</span>  </div>  <div>Keywords:  <span itemprop="keywords">Proteomics</span>  </div>  <div>Version:  <span itemprop="softwareVersion">0.4.1</span>  </div>  ...  </div> |

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 Tools properties as RDFa within HTML** |
| <div vocab="http://schema.org/" typeof="SoftwareApplication">  <div property="name">Pladipus</div>  <div property="description">The [CompOmics](http://www.compomics.com/) Pladipus system is an affordable cluster computing framework allowing for distributed processing, management and analyses for bio-informatics pipelines.</div>  <div>Application categories:  <span property="applicationCategory">Command line</span>,  <span property="applicationCategory">Desktop GUI</span>  </div>  <div>Keywords:  <span property="keywords">Proteomics</span>  </div>  <div>Version:  <span property="softwareVersion">0.4.1</span>  </div>  ...  </div> |

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

# 

# 

# 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

|  |
| --- |
| **Example 6. Extended example using JSON-LD, RDFa and microdata** |
| *JSON-LD:*  <script type="application/ld+json">  {  "@context": "http://schema.org",  "@type": "SoftwareApplication",  "name": "SNPs and GO",  "description": "SNPs&GO is a server for the prediction of single point protein mutations likely to be involved in the insurgence of diseases in humans.",  "url": "http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html",  "applicationCategory": "Tool",  "Keywords": "DNA polymorphism",  "Keywords": "Bioinformatics",  "Keywords": "Protein properties",  "softwareVersion": "1.0",  "featureList": "Variant classification",  "publisher": {  "@type": "Person",  "name": "Rita Casadio",  "email": "casadio@biocomp.unibo.it"  },  "publisher": {  "@type": "Person",  "name": "Pier Luigi Martelli",  "email": "gigi@biocomp.unibo.it"  },  "operatingSystem": "All"  }  </script>  *RDFa:*  <div vocab="http://schema.org/" typeof="SoftwareApplication">  <div property="name">SNPs and GO</div>  <div property="description">SNPs&GO is a server for the prediction of single point protein mutations likely to be involved in the insurgence of diseases in humans.</div>  <div>Url:  <a property="url" href="http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html"> http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html </a>  </div>  <div>Type:  <span property="applicationCategory">Tool</span>  </div>  <div>Keywords:  <span property="keywords">DNA polymorphism</span>,  <span property="keywords">Bioinformatics</span>,  <span property="keywords">Protein properties</span>  </div>  <div>Version:  <span property="softwareVersion">1.0</span>  </div>  <div>Function:  <span property="featureList">Variant classification</span>  </div>  <div>Contacts:  <div property="publisher">  <div typeof="<http://schema.org/Person>">  <div property="name">Rita Casadio</div>  <div property="email">casadio@biocomp.unibo.it</div>  </div>  <div typeof="<http://schema.org/Person>">  <div property="name">Pier Luigi Martelli</div>  <div property="email">gigi@biocomp.unibo.it</div>  </div>  </div>  </div>  <!-- The following is to appease Google's Structured Data  Testing Tool, which requires at least two properties from these four: operatingSystem, aggregateRating, applicationCategory, offers. -->  <meta property="operatingSystem" content="All" />  </div>  *Microdata:*  <div itemscope itemtype="http://schema.org/SoftwareApplication">  <div itemprop="name">SNPs and GO</div>  <div itemprop="description">SNPs&GO is a server for the prediction of single point protein mutations likely to be involved in the insurgence of diseases in humans.</div>  <div>Url:  <a itemprop="url" href="http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html"> http://snps-and-go.biocomp.unibo.it/snps-and-go/index.html </a>  </div>  <div>Type:  <span itemprop="applicationCategory">Tool</span>  </div>  <div>Keywords:  <span itemprop="keywords">DNA polymorphism</span>,  <span itemprop="keywords">Bioinformatics</span>,  <span itemprop="keywords">Protein properties</span>  </div>  <div>Version:  <span itemprop="softwareVersion">1.0</span>  </div>  <div>Function:  <span itemprop="featureList">Variant classification</span>  </div>  <div>Contacts:  <div itemprop="publisher">  <div itemscope itemtype="<http://schema.org/Person>">  <div itemprop="name">Rita Casadio</div>  <div itemprop="email">casadio@biocomp.unibo.it</div>  </div>  <div itemscope itemtype="<http://schema.org/Person>">  <div itemprop="name">Pier Luigi Martelli</div>  <div itemprop="email">gigi@biocomp.unibo.it</div>  </div>  </div>  </div>  <!-- The following is to appease Google's Structured Data  Testing Tool, which requires at least two properties from these four: operatingSystem, aggregateRating, applicationCategory, offers. -->  <meta itemprop="operatingSystem" content="All" />  </div> |