
Tool-Information-Standard Documentation

Release latest

May 16, 2018

Contents

1	Why is a standard needed at all?	3
2	Description of attributes	5
2.1	General attributes	7
2.2	“Documentation” group	8
2.3	“Code availability” group	8
2.4	“Accessibility” group	8
2.5	“Support” group	9
2.6	“Downloads” group	10

This is the documentation for the [Tool Information Standard](#).

Contents:

Tool Information Standard An information standard for the description of bioinformatics software tools.

The standard described below represents a “respectable beta” and will be implemented in <https://bio.tools>. Please join the [discussion on GitHub](#)

[biotoolsSchema](#) provides the foundation for an *information standard* for the description of tools that is being adopted by <https://bio.tools>. The standard has two components:

- a list of tool attributes (see Table below) that must be specified (or stated as not being available) for an entry to be assigned in a 5 tier ratings of entry completeness and quality. Some attributes are grouped (see tables on right) for purposes of determining adherence to the standard.
- a set of [Curation Guidelines](#) describing *how* each attribute should be specified, *i.e.* concerning the *quality* of an entry. The guidelines go beyond the syntactic and semantic constraints defined by [biotoolsSchema](#).

Adherence to some of guidelines is verified by automated QC checks within bio.tools, whereas others must be verified by manual inspection.

The standard provides a basis for monitoring of content and labelling of bio.tools entries in various ways:

- *entry completeness* (“SPARSE” through to “COMPREHENSIVE”)
- whether an entry was manually inspected
- whether an entry is verified, *i.e.* conforms to the guidelines, passing both the automated and manual checks (stamps of dates of verification and last update would be needed)
- overall *entry quality* (from “NEEDS TO IMPROVE” through to “EXCELLENT”) conflating notions above (see infographic below)

If and how to use labels is currently [under discussion](#).

The standard will be applied to bio.tool as follows:

- “SPARSE” : minimum information requirement for new entries. Such entries will be invisible by default (see below) and this made clear to the registrant via the UI.
- “BASIC DETAILS” : level at which an entry becomes visible (by default). Entry visibility will be settable by the user with an option to “Show more results” (*i.e.* over invisible entries)

	SPARSE	BASIC DETAILS	DETAILED	HIGHLY DETAILED	COMPREHENSIVE	
						→ Documentation General Manual API documentation API specification At least one
Name	✓	✓	✓	✓	✓	
Description	✓	✓	✓	✓	✓	
Homepage	✓	✓	✓	✓	✓	
Unique ID	✓	✓	✓	✓	✓	
Tool type		✓	✓	✓	✓	→ Code availability Repository Source code Source package At least one
Scientific topic(s)		✓	✓	✓	✓	
Publication (1)		✓	✓	✓	✓	
Support → (2)		✓	✓	✓	✓	→ Accessibility Terms of use Accessibility Cost At least one
Scientific operation(s)			✓	✓	✓	
Documentation (3) →			✓	✓	✓	
Operating system			✓	✓	✓	→ Support Helpdesk Issue tracker Mailing list Contact person At least one
Language			✓	✓	✓	
License (4)			✓	✓	✓	
Type of input & output data (5)				✓	✓	
Accessibility →				✓	✓	
Code availability (3) →				✓	✓	
Downloads (3) →				✓	✓	
Supported data format(s) (6)					✓	→ Downloads Biological data Binaries Binary package Container file CWL file Ontology VM image Tool wrapper (galaxy) Tool wrapper (taverna) Tool wrapper (other) At least one
Scientific benchmark (3)					✓	
Technical monitoring (3)					✓	

- (1) A valid identifier (DOI, PMID or PMCID) or "Unpublished" must be specified.
(2) An appropriate link or email annotated as the "primary contact" must be specified.
(3) An appropriate link must be specified.
(4) Either a valid license or "Unlicensed" must be specified.
(5) At least one input or output must be specified.
(6) At least one data format for each input and output must be specified.

	NEEDS TO IMPROVE	OKAY	GOOD	VERY GOOD	EXCELLENT
	 The entry lacks basic information and may not conform to bio.tools curation guidelines. Improvements are needed.	 The entry has basic information but lacks detailed scientific annotations, and may not conform to bio.tools curation guidelines.	 The entry is detailed but would benefit from richer description e.g. tool inputs & outputs. It may not conform to bio.tools curation guidelines.	 The entry is nearly complete and conforms to bio.tools curation guidelines. More details e.g. data formats can be added.	 The entry is comprehensive and conforms to bio.tools curation guidelines. Further improvement may be possible.
Complete?	SPARSE	BASIC DETAILS	DETAILED	HIGHLY DETAILED	COMPREHENSIVE
Valid syntax? biotoolsSchema 3.0.0 compatible	✓	✓	✓	✓	✓
Manually inspected?	?	?	✓	✓	✓
Conforms to guidelines? bio.tools Curation Guidelines 1.0	?	?	?	✓	✓

Why is a standard needed at all?

There are several reasons why a community-defined standard for tool information is helpful:

- accessible summary of what type of information bio.tools provides
- flexible information requirement, including a minimum requirements presenting a low barrier to new registrations, compatible / enabling integration with other major related cataloguing efforts (e.g. BioContainers)
- quality tiers motivate individual entry owners to improve their entries (curation as a “game”), with a “gold standard” of entry quality for curators to aspire to
- a framework / workflow to guide tasks and priorities of curators, thematic editors and bio.tools admin
- a basis for metrics of bio.tools quality, KPIs (key performance indicators) of quality improvement objectives, and targets
- a component of branding bio.tools as a trusted source of quality tool information

CHAPTER 2

Description of attributes

Attributes defined as mandatory in the latest stable schema ([biotools-2.0.0](#)) are those in the “OKAY” tier plus “Scientific operation” from the “GOOD” tier.

2.1 General attributes

Attribute	Description	Format	element
Name	Canonical software name assigned by the software developer or service provider	Text	<name>
Description	Short and concise textual description of the software function	Text	<description>
Homepage	Homepage of the software, or some URL that best serves this purpose	URL	<homepage>
Unique ID	Unique ID of the tool that is assigned upon registration of the software in bio.tools	Text (URL-safe version of tool name)	<toolID>
Tool type	The type of application software: a discrete software entity can have more than one type.	enum (from biotoolsS-chema, see below)	<toolType>
Scientific topics	General scientific domain the software serves or other general category, <i>e.g.</i> ‘Proteomics’	Term and / or URI of EDAM Topic concept(s) (1)	<topic>
Publications	Publications about the software	DOI, PMID or PM-CID	<publication>
Scientific operations	The basic operation(s) performed by the software, <i>e.g.</i> ‘Multiple sequence alignment’	Term and / or URI of EDAM Operation concept(s)	<function><operation>
Operating system	The operating system supported by a downloadable software package.	enum (from biotoolsS-chema)	<labels><Operating system>
Language	Name of programming language the software source code was written in, <i>e.g.</i> ‘C’.		<language>
License	Software or data usage license	enum (from biotoolsS-chema)	<labels><license>
Type of input & output data	Type of primary input / output data (if any), <i>e.g.</i> ‘Protein sequences’	Term and / or URI of EDAM Data	<function><input>/<output><data>
2.1. General attributes		concept(s)	7
Supported data formats	Allowed format(s) of primary inputs/outputs, <i>e.g.</i> ‘FASTA’	Term and / or URI of	<function><input>/<output><format>

2.2 “Documentation” group

Attribute	Description	Format	element
General documentation	General documentation	URL	<documentation><type>
Manual	Information on how to use the software.	URL	<documentation><type>
API documentation	Human-readable API documentation.	URL	<documentation><type> <documentation>/<type>
API specification	File providing an API specification for the software, e.g. Swagger/OpenAPI, WSDL or RAML file.	URL	<download><type>API specification</type>

2.3 “Code availability” group

Attribute	Description	Format	element
Repository	Link to repository where source code, data and other files may be downloaded	URL	<link><type>Repository</type>
Source code	Software source code.	URL	<download><type>Source code</type>
Source package	Source package (of various types) for the software.	URL	<download><type>Source package</type>

2.4 “Accessibility” group

Attribute	Description	Format	element
Terms of use	Rules that one must agree to abide by in order to use a service.	URL	<link><type>Terms of use</type>
Accessibility	Whether the software is freely available for use.	enum (from biotoolsSchema)	<labels><Accessibility>
Cost	Monetary cost of acquiring the software.	enum (from biotoolsSchema)	<labels><Cost>

2.5 “Support” group

Attribute	Description	Format	element
Helpdesk	Helpdesk providing support in using the software.	URL	<link><type>Helpde type>
Issue tracker	Link to tracker for software issues, bug reports, feature requests etc.	URL	<link><type>Issue tracker</ type>
Mailing list	Link to mailing list for software announcements, discussions, support etc.	URL	<link><type>Mailin list</ type>
Contact person	Primary contact, <i>e.g.</i> a person, helpdesk or mailing list	Name, email, URL and/or OR- CID iD	<credit><typeRole> contact</ typeRole>

2.6 “Downloads” group

Attribute	Description	Format	element
Biological data	Biological data, or a web page on a database portal where such data may be downloaded.	URL	<code><download><type>Biological data</type></code>
Binaries	Binaries for the software.	URL	<code><download><type>Binaries</type></code>
Binary package	Binary package for the software.	URL	<code><download><type>Binary package</type></code>
Container file	Container file including the software.	URL	<code><download><type>Container file</type></code>
CWL file	Common Workflow Language (CWL) file for the software.	URL	<code><download><type>CWL file</type></code>
Ontology	A file containing an ontology, controlled vocabulary, terminology etc.	URL	<code><download><type>Ontology</type></code>
VM image	Virtual machine (VM) image for the software.	URL	<code><download><type>VM image</type></code>
Tool wrapper (galaxy)	Galaxy tool configuration file (wrapper) for the software.	URL	<code><download><type>Tool wrapper (galaxy)</type></code>
Tool wrapper (tav-erna)	Taverna configuration file for the software.	URL	<code><download><type>Tool wrapper (taverna)</type></code>
Tool wrapper (other)	Workbench configuration file (other than taverna, galaxy or CWL wrapper) for the software.	URL	<code><download><type>Tool wrapper (other)</type></code>

Note:

1. EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics, including types of data and data identifiers, data formats, operations and topics. EDAM provides a set of terms with synonyms and definitions - organised into an intuitive hierarchy for convenient use. You can find [EDAM on GitHub](#).