**Metrics Framework Infrastructure Specification**

**Abstract**

This document defines the Metrics Framework Infrastructure data format for registering and monitoring metrics and collecting data from metrics executions.

**Status of This Document**

This is an editor's draft of a document for the Development and Analysis of ELIXIR Metrics Framework. Distribution of this document is unlimited.

**Table of Contents**

[Introduction](#h.5xxxgni0z5v5)

[Basic Grammar and Conventions](#h.pl1vjm5azbaq)

[Metrics Framework Definition](#h.dfjx47pj1nnm)

[Elements](#h.m196e28lm3e1)

[Component](#h.5zsmmgv7v4gz)

[Components Registry](#h.2cljx6zjbc2)

[Data and Monitoring Repository](#h.caw0a85felrb)

[Metrics Framework Functionality](#h.pi4ibj989bxr)

[Component Description](#h.902n5rg4u5ll)

[Component Registry](#h.osgsqwhryh8j)

[Store and Monitoring Data](#h.cxdq9qq29sr9)

[Metrics Framework Analysis and Design](#h.3295wc23400s)

[Example](#h.5jpu7fq4zulf)

[References](#h.ibfbiqpnunew)

# **Introduction**

Measuring the impact of bioinformatics resources is of paramount importance. Metrics encourage attainment of certain standards of quality of resources. We propose the development of a metrics framework to provide an infrastructure capable of collect/register different metrics definitions and their functionalities, process them and provide a user interface to access these results. The data collect of each metric can be used as an indicator of quality that measure the impact of specific bioinformatic resource (software, biological databases, knowledge diffusion, computational performance, etc.) which suggesting a certain standard of quality.

This framework is formed by three elements which work together for carrying out the complete registry and monitoring processes. The first element Component, define the metric schema and functionality ; the second component Components Registry, register the component metadata into a registry to make available for use; The third element Data and Monitoring Repository, install and execute components and collect data from the component execution. Data generated from each component execution will be able for their visualization

This document provides the Metrics Framework specification, in addition to definition, description and functionality of each element of this infrastructure. Also, this document contains the system analysis and design.

# **Basic Grammar and Conventions**

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](https://tools.ietf.org/html/rfc2119)].

# Metrics Framework Definition

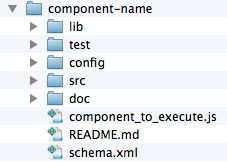
## Elements

The metrics framework design follows a modular structure, consisting of three autonomous elements: Component, Components Registry and, Data and Monitoring Repository. These elements interact with each other and beside this interaction, one user can also interact independently with each element using an user interface adjust to each setting.

### Component

A component is the basic unit defined into the Metrics Framework. Each component are defined through two differentiated parts:

* ***Metric standard definition***, following a common schema. This schema is described in a metadata file that defines a set of descriptors used in deploying a component. Examples of descriptors are: name, dependencies, frequency, resource, output, repository, etc.
* ***Metric functionality***, following a common structure. Code, documentation, guidelines, and examples can define this structure.

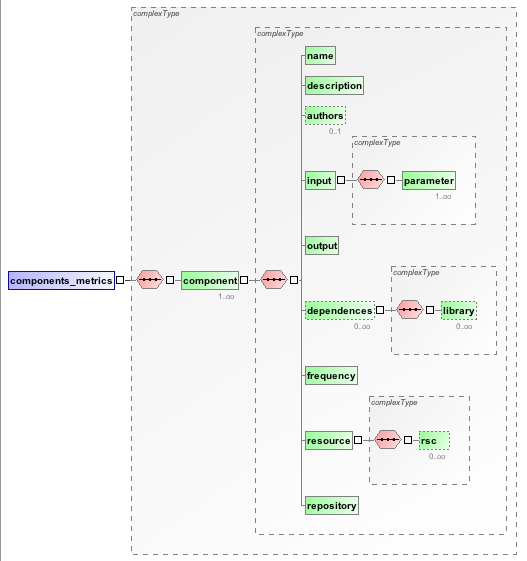


Due to the distributed approach of this framework, the components must be stored in a Source Code Manager (e.g. GitHub) that allows an easy and quick access to each component.

The basic component structure suggested is shown in Figure 1.

*Figure 1. Basic component structure*. The schema.xml file describes the metadata information or the Metric standard definition. Other directories and files define the metric functionality (Code, documentation, guidelines, and examples)

The file schema.xml describes the component metadata. This metadata contains essential information that will be used to register, execute and collect data from this component. The Figure 2 shows the schema architecture and the meanings of each parameter are given in the Table 1.



*Figure 2. Graphical component schema.* This graphic shows the component schema hierarchy.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | **Description** | **Cardinality** | **Type** |
| **Name** | Component name | 1 | String |
| **Description** | Short component explanation | 1 | String |
| **Authors** | Authors names separated by commas | 0..1 | String |
| **Input** | Input parameters list used in the component execution. | 1..\* | String |
| **Output** | Type execution result: number, string or list. | 1 | String |
| **Dependencies** | Libraries list necessaries for this component. | 0..\* | String List |
| **Frequency** | Time interval to execute a component: monthly, weekly, daily. | 1 | String |
| **Resources** | Resource list representation: database, software, training, computational. | 1..\* | String List |
| **Repository** | Source Manger Code URL where the component is stored. | 1 | String |

*Table 1. Component schema.* This table describes in detail, the parameters which defines the components metadata.

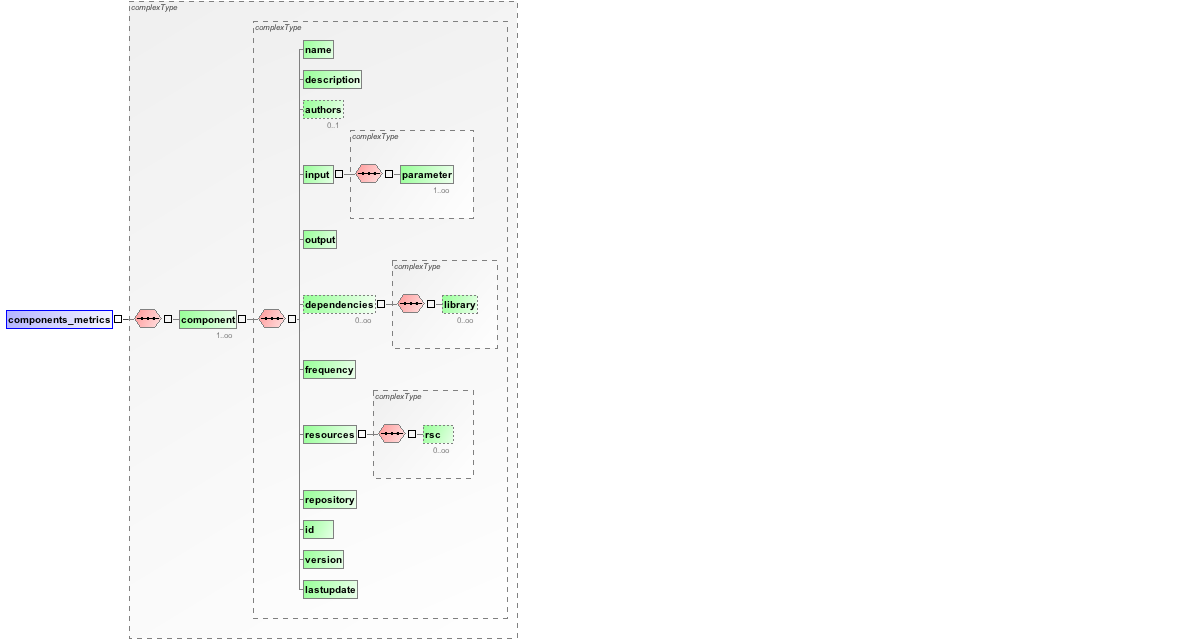
### Components Registry

The Components Registry contains metadata information of every component that has been registered. This information will be used when a specific component was installed.

The metadata information is described in a common schema consisting of component metadata (described in section 3.1.1) and additional descriptors required for registering a component. When a component is registered must contain this schema definition in its metadata file. This schema configuration is shown in Figure 3, and theirs parameters definition are described in Table 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | **Description** | **Cardinality** | **Type** |
| **Name** | Component name | 1 | String |
| **Description** | Short component explanation | 1 | String |
| **Authors** | Authors names separated by commas | 0..1 | String |
| **Input** | Input parameters list used in the component execution. | 1..\* | String |
| **Output** | Type execution result: number, string or list. | 1 | String |
| **Dependencies** | Libraries list necessaries for this component. | 0..\* | String List |
| **Frequency** | Time interval to execute a component: monthly, weekly, daily. | 1 | String |
| **Resources** | Resource list representation: database, software, training, computational. | 1..\* | String List |
| **Repository** | Source Manger Code URL where the component is stored. | 1 | String |
| **Id** | Unique identify, this is composed for the component name and one exclusive alphanumeric. | 1 | String |
| **version** | Component version. | 1 | String |
| **lastupdated** | Component last date of upgrade. | 1 | Date |
| **...** |  |  |  |

*Table 2. Component Registry schema.* This table describes in detail, the parameters which defines the metadata of components registered. The orange cells show additionals descriptors as compared to component metadata schema.



*Figure 3. Graphical Component Registry schema.* This graphic shows the component schema hierarchy consisting of essential component metadata and additional descriptors: id, version, lastupdate.

### Data and Monitoring Repository

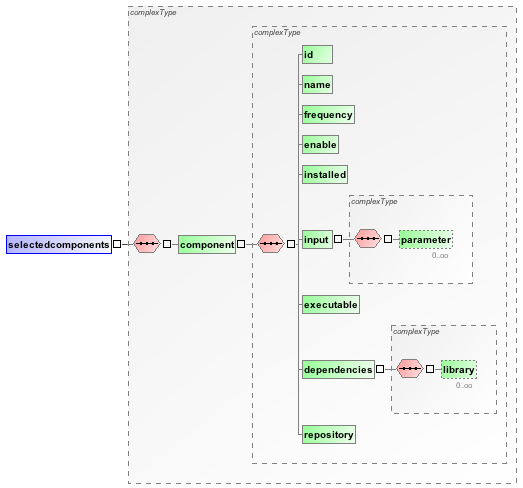
This Repository collects data coming from a set of components registered into the Components Repository; also it makes available these data for different users.

As a part of its configuration, the Data and Monitoring Repository makes use of a schema from where, it extracts necessary information to install and execute a specific component.

This schema configuration is shown in Figure 4, and theirs parameters definition are described in Table 3.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameters** | **Description** | **Cardinality** | **Type** |
| **Id** | Unique identify, this is composed for the component name and one exclusive number. | 1 | String |
| **Name** | Component name. | 1 | String |
| **Frequency** | Time interval to execute a component: monthly, weekly, daily. | 1 | String |
| **Enable** | Define if the component functionality is available into the source code manager. | 1 | Boolean |
| **Installed** | Define if the component is installed. | 1 | Boolean |
| **Input** | Parameter list for executing a component. | 0..\* | String |
| **Executable** | File name of the executable component. | 1 | String |
| **Dependencies** | Libraries list necessaries for this component. | 0..\* | String  List |
| **Repository** | Source Manger Code URL where the component is stored. | 1 | String |

*Table 3. Selected Component schema.* This table describes in detail, the parameters which def



*Figure 4. Graphical Data and Monitoring schema.* This graphic shows the selected components schema hierarchy.

## Metrics Framework Functionality

### Component Description

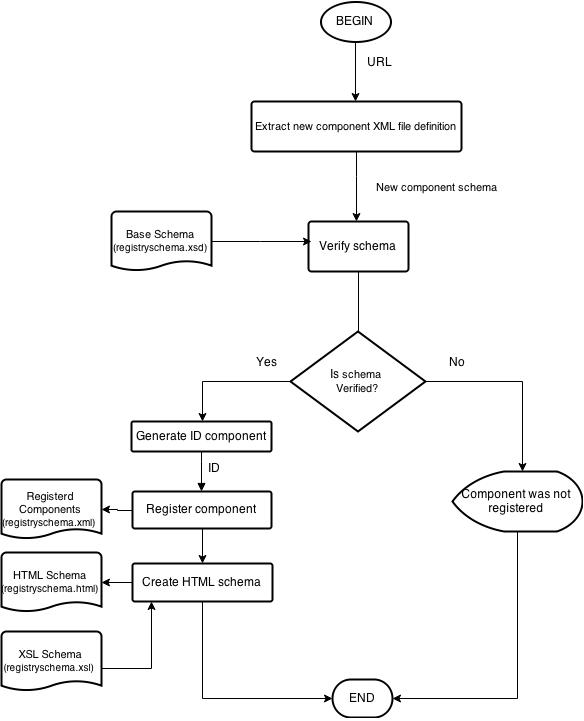
A component should describe its metadata information in a XML file, following the specification explained in the section 3.1.1. The schema of this file is described in the Table 1 and the Figure 2.

The component implementation requires a based development in NodeJS [NodeJS] technology however, a component can be registered although its implementation is not even defined. This component can only be installed and executed when its implementation has been include into the Source Manager Content where is contained.

### Component Registry

The registry component follows the next actions (Figure 5):

1. New component must provide the URL where the component metadata is stored (XML file).
2. The XML definition is extracted from this URL.
3. The new component XML schema is compared with the base schema defined into the Component Registry.
4. If this schema is verified, the unique component ID is generated. The final ID will be composed of the component name and an unique alphanumeric string (**name-alphanumeric\_string**).
5. The new component metadata definition, and it ID, is added in the registered components XML file,
6. Finally, the HTML schema is created, adding the new component.



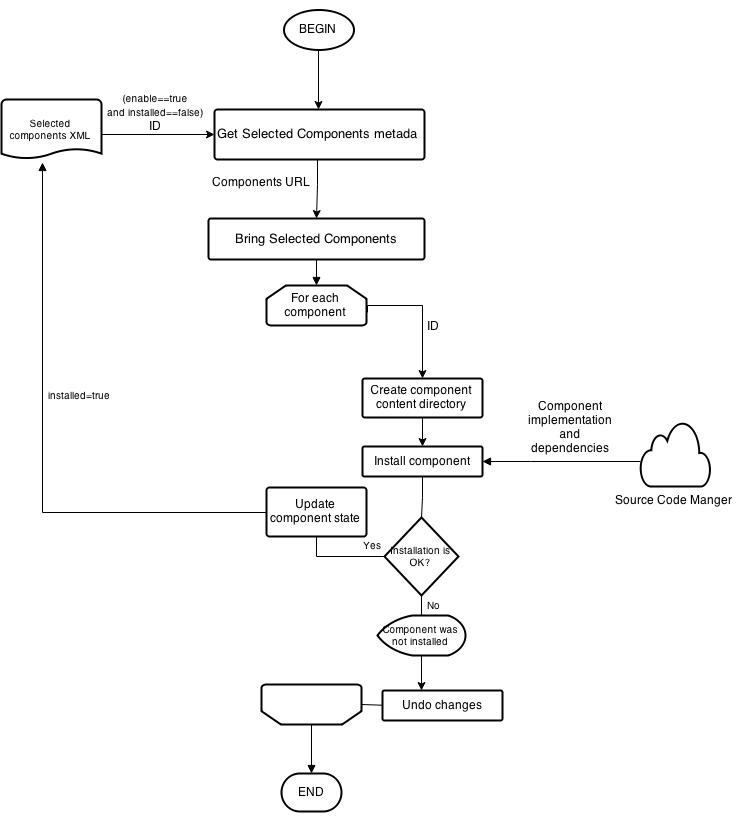
*Figure 5. Registry Component Flowchart.* This graphic shows the procedure to register a new component.

### Store and Monitoring Data

The Data and Monitoring Repository install and execute a specific group of components. Moreover, the data from these executions are stored in this repository and then, different users can access to this repository visualizing particular information.

These are the main tasks performed in this repository:

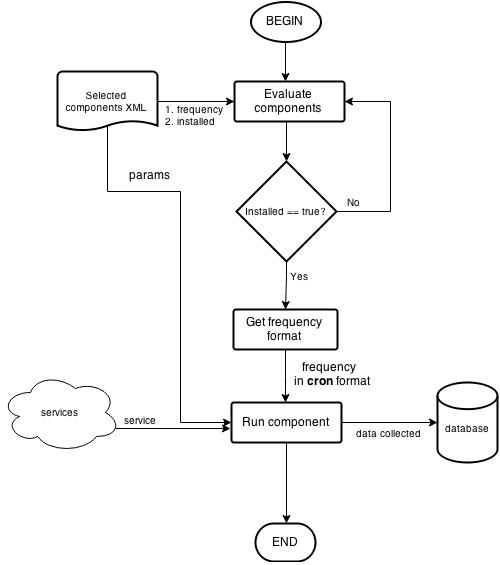
1. Install components (Figure 6):
   1. Evaluate the list of Selected Components (XML file) to install a component.
   2. Get the URL from each ***not*** *installed* component with *enable* parameter equal ***true***.
   3. Access to Source Code Manager (GitHub) and bringing the component implementation.
   4. Create a directory component to store the component implementation. This directory will be called as the component ID.
   5. Bring the component implementation.
   6. Install component (and its dependencies).
   7. If the installation is successful, updated the component ID in the Selected Components XML schema.
   8. If not, undo the changes



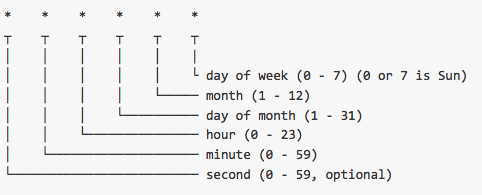
*Figure 6. Install Component Flowchart.* This graphic shows the procedure to install a set of selected component.

1. Run components (Figure 7):
   1. Evaluate the list of Selected Components (XML file) to run a component.
   2. Select each component with ***installed*** descriptor equal **true**.
   3. Get frequency descriptor format: supported format **cron-parser** (Figure 8)
   4. Run each component according its execution time interval (frequence):
      * Execution format: ***node executable [param1 param2 … paramN]***, where:

* **executable**: is the filename of the executable component (see Table 2).
* **[param1 param2 … paramN]**: params of execution (see Table 2).
  1. Collect the data from every component execution.



*Figure 7. Run Component Flowchart.* This graphic shows the procedure to execute a set of selected component.

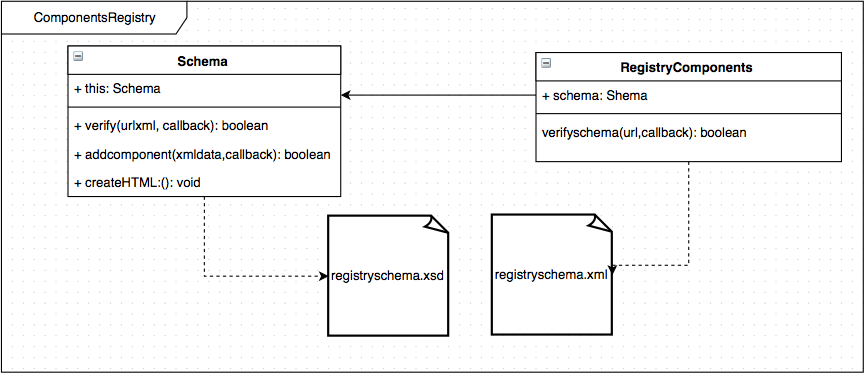


# 

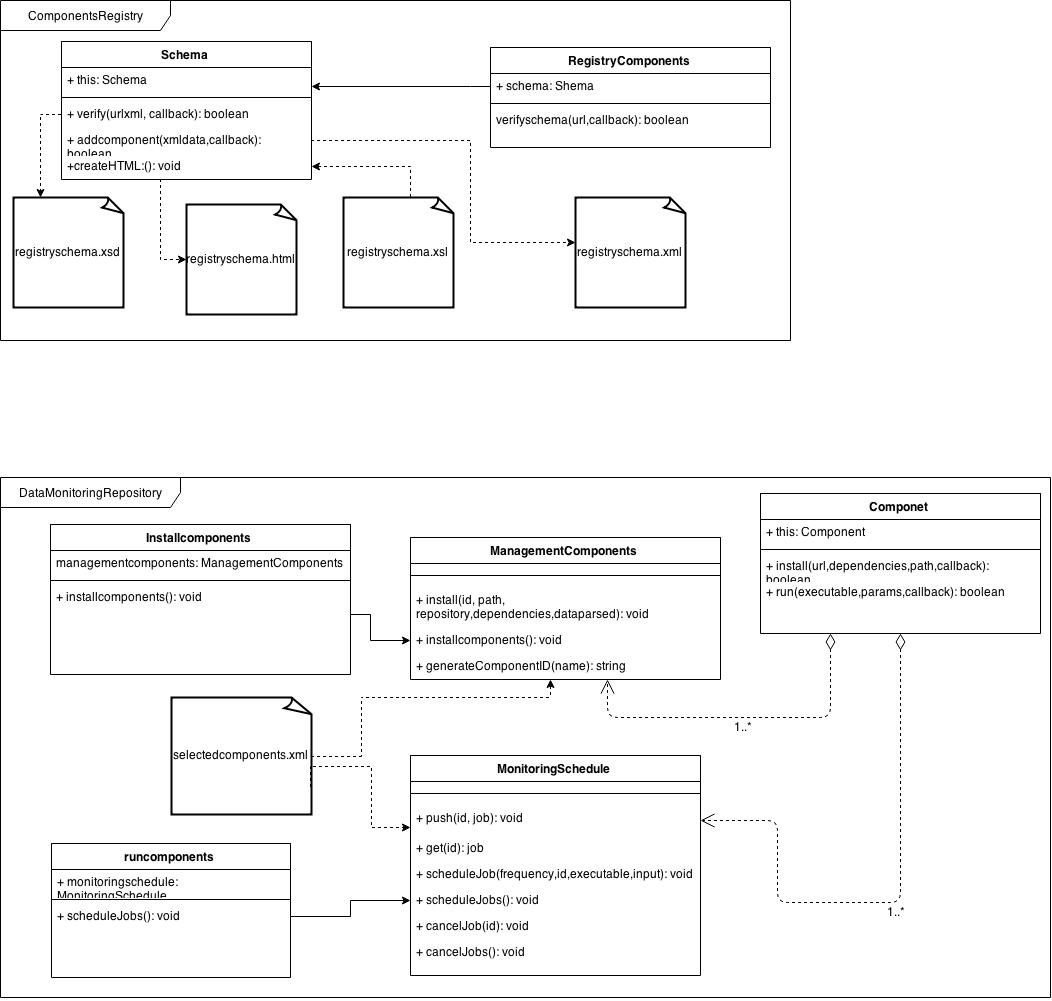
*Figure 8. Cron-parse format*. This format is used to get the execution time interval of one component .

1. Collects data from the components execution. These data are stored into the repository and made available via UI (API, WS or GUI) (It needs more detail).

# Metrics Framework Analysis and Design



*Figure 9. ComponentsRegistry Module. Classes and methods responsible for the component registry compose this module*



*Figure 10. DataMonitoringRepository Module. Classes and methods responsible for the component installation and execution compose this module.*

# Example

This section gives an example concerning the definition, registry, monitoring of a Citation Metric-Component.

**Citation Metric**: defines a resource search in papers published in Pubmed Central repository. It includes all words and numbers in the title, abstract and article body, as well as in table and figure captions and in the article reference section.

This metric can be used to provide evidence of the impact of the corresponding resource (e.g. database).

**Citation Component**: this describes the Citation Metric definition (metadata) in a XML schema, and also defines the Citation Metric implementation. This implementation (code, documentation, examples, etc.) is stored into Source Code Manager (e.g. Github).

|  |
| --- |
| component.png |

**Component Schema**: This metadata is used for registering the component. The main information for future execution are defined by:

* Name: used to generate the ID component.
* Input: resource search, in this example a database Uniprot.
* Dependences: libraries necessary to run this component.
* Resource: kind of resource, in this example a database Uniprot.
* Repository: URL where the implementation of this component is stored.

|  |
| --- |
| <?xml version="1.0" encoding="utf-8"?>  <components\_metrics>  <component>  <name>metric-module-citations</name>  <description>Citations in publications</description>  <authors>names of authors</authors>  <input>  <parameter>swissprot,swiss prot,swiss-prot,UniProtKB,uniprot</parameter>  </input>  <output>number</output>  <dependences>  <library>xml2js</library>  </dependences>  <frequency>monthly</frequency>  <resource>  <rsc>database</rsc>  </resource>  <repository>https://github.com/elixirhub/metrics-framework.git</repository>  </component>  </components\_metrics> |

**Component Implementation**: composed of code, documentation, examples, guidelines, and others, which defined the component functionality. For this example the Component Citation is stored into Github source code manager. The code implementation must be distributed in **library** and **test** folder. The metadata schema is described in **schema.xml** file.

|  |
| --- |
|  |

**Component Registry Schema**: metadata from each registered component plus extra descriptors for management of component registry. Additional information are defined in:

* Id: Component ID (name+unique\_alphanumeric).
* Version: alpha, beta, final, etc.
* Lastupdated: date of last upgrade.

This schema is adapted to increase the descriptors according to the model needs.

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8" standalone="yes"?>  <components\_metrics>  <component>  <name>metrics-module-citation</name>  <description>Citations in publications</description>  <authors>names of authors</authors>  <input>  <parameter>swissprot,swiss prot,swiss-prot,UniProtKB,uniprot</parameter>  </input>  <output>number</output>  <dependencies>  <library>xml2js</library>  </dependencies>  <frequency>monthly</frequency>  <resources>  <rsc>database</rsc>  </resources>  <repository>https://github.com/elixirhub/metrics-module-citation.git</repository>  <id>metrics-module-citation-EJn891E8</id>  <version>final</version>  <lastupdate>12/5/2015</lastupdate>  </component>  </components\_metrics> |

**Data and Monitoring Schema**: metadata of selected component with extra descriptors for management of monitoring and data collecting. Additional information are defined in:

* Enable: the implementation code is available into source code manager.
* Installed: this component and its dependencies have been installed therefore it is ready to run.

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8" standalone="yes"?>  <selectedcomponents>  <component>  <id>metrics-module-citation-4ywN\_j5H</id>  <name>metrics-module-citation</name>  <frequency>minute</frequency>  <enable>true</enable>  <installed>false</installed>  <input>  <parameter>swissprot,swiss prot,swiss-prot,UniProtKB,uniprot</parameter>  </input>  <executable>test/testcitations.js</executable>  <dependencies>  <library>xml2js</library>  </dependencies>  <repository>https://github.com/elixirhub/metrics-module-citation.git</repository>  </component>  </selectedcomponents> |

**Install execution**: example of console log for the component installation.The parameters of execution are defined in XML schema:

* Component ID: metrics-module-citation-4ywN\_j5H
* Enable: true (implementation is available).
* Parameter: swissprot,swiss prot,swiss-prot,UniProtKB,uniprot (resource)
* Dependencies: xml2js (libraries necessaries for this component).
* Repository: <https://github.com/elixirhub/metrics-module-citation.git> (URL from where extracts the component implementation)

|  |
| --- |
| [15:06:23][admin]~/DataMonitoringRepository$ node installcomponents.js  https://nodeload.github.com/elixirhub/metrics-module-citation/zip/master  https://github.com/elixirhub/metrics-module-citation.git download into components/metrics-module-citation-4ywN\_j5H completed  installing library xml2js  xml2js@0.4.9 ../node\_modules/xml2js  ├── sax@0.6.1  └── xmlbuilder@2.6.4 (lodash@3.9.3)  Library xml2js installed  Component metrics-module-citation-4ywN\_j5H has been updated in schema/selectedcomponents.xml |

**Monitoring execution**: example of console log for the component monitoring. The parameters of execution are defined in XML schema:

* Component ID: metrics-module-citation-4ywN\_j5H
* Frequency: minute
* Installed: true (upgrade in the installation execution).
* Parameter: swissprot,swiss prot,swiss-prot,UniProtKB,uniprot
* Executable: test/testcitations.js

|  |
| --- |
| [15:24:33][admin]~/DataMonitoringRepository$ node runcomponents.js  components/metrics-module-citation-4ywN\_j5H/test/testcitations.js  Frequency: minute (cron format 0 \* \* \* \* \*)  Resource: "swissprot,swiss prot,swiss-prot,UniProtKB,uniprot"  30411 citations for swissprot,swiss prot,swiss-prot,UniProtKB,uniprot  data save into > citation.txt  component metrics-module-citation-4ywN\_j5H was executed |

# References

[NodeJS]

<https://nodejs.org/>

[RFC2119]

Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](https://tools.ietf.org/html/bcp14), [RFC 2119](https://tools.ietf.org/html/rfc2119), March 1997.