MONGODB BACKEND SCHEMA.

Tools Monitoring mongodb database has four collection:

- **tools** – collection that contains all tool's information collected from various repositories.
- **metrics** - collection that contain tools' metrics mostly obtained from **tools** data.
- **tools.log** – a history of tools collection modifications.
- **metrics.log** – a history of metrics collection modifications.

The structure of tools collection resembles the json schema provided by the tools monitoring REST API with difference in mongodb primary key which is hidden from the end client.

TOOLS

Tools Monitoring REST API exposes tools' identifiers in a way of URL. For instance:

https://openebench.bsc.es/monitor/tool/biotools:pmut:2017/rest/mmb.irbbarcelona.org

is stored in mongodb as:

```
"_id": {
        "id": "pmut",
        "nmsp": "biotools",
        "version": "2017",
        "type": "rest",
        "host": "mmb.irbbarcelona.org"
}
```

For clients, the REST API exposes synthetic fields:

```
"@id": "https://openebench.bsc.es/monitor/tool/biotools:pmut:2017/rest/mmb.irbbarcelona.org",
"@label": "pmut",
"@version": "2017",
"@type": "rest",
"@timestamp": "2018-08-30T09:32:23.446Z",
"@license": "https://creativecommons.org/licenses/by/4.0/"
```

Note that **@id**, **@type** and **@label** properties are fully compatible with JSON-LD.

Identifier parts description:

property	description
id	Principal tool identifier. This should be an original tool name (i.e. "blast").
nmsp	The origin of the record. Tools data may come from different sources/repositories (i.e. "biotools", "bioconda", "galaxy", "github".
version	Tool's version. There could be several versions of tools (i.e. "1.3.1", "2017", etc.)
type	Tools may be implemented or accessed in different ways via web interface, REST API, command line, workflow, etc. Current types are ("cmd", "web", "db", "app", "lib", "ontology", "workflow", "plugin", "sparql", "soap", "script", "rest", "workbench", "suite")
host	This is a tool's authority. There could be various deployments of the same tool in different institutions. The host name is used for this purpose.

All documents with a same id, independently of their provenance, should refer to the same tool:

biotools:**mytool**:2017/rest/my.organization.org
bioconda:**mytool**:1.3/cmd/other.organization.org

When original tool's identifier is different to the one used by Tools Monitoring, it is stored in the "external identifier" or "**xid"** property:

"@id": "https://dev-openebench.bsc.es/monitor/tool/biotools:**trimal**:1.4/cmd/trimal.cgenomics.org"
"**xid**": "**trimAl**:1.4"

There is a special record that contains only a principal identifier:

https://openebench.bsc.es/monitor/tool/pmut

This record (with other primary key fields set to 'null' value) is used to keep a global tool descriptions.

METRICS

The **@id** for the metrics follows the same scheme:

https://openebench.bsc.es/monitor/metrics/biotools:pmut:2017/rest/mmb.irbbarcelona.org

Unlike for the tools, the primary key "**_id**" for the metrics collection is not compound:

"_id": "biotools:pmut:2017/rest/mmb.irbbarcelona.org"

There are also less synthetic properties:

```
"@id": "https://openebench.bsc.es/monitor/metrics/biotools:pmut:2017/rest/mmb.irbbarcelona.org",
"@type": "metrics",
"@timestamp": "2018-08-31T01:16:19.644Z",
"@license": "https://creativecommons.org/licenses/by/4.0/
```

LOGS

Both, "**tools**" and "**metrics**" collections have their log counterparts: "**tools.log**" and "**metrics.log**".

```
"_id": {
        "id": "biotools:oases_cloudifb:0.2.08/cmd/www.ebi.ac.uk",
        "date": "2017-12-11T12:14:40.153Z"
},
"src": "biotools",
"patch": []
```

Log collections have the same structure and store corresponding modifications in JSON Patch format (rfc6902). Any log record has a timestamp and a source of the update (the user that made the modifications).