

djehuty

The 4TU.ResearchData repository system



<https://github.com/4TUResearchData/djehuty>
version 0.0.1, March 20, 2023

Contents

1	Introduction	1
1.1	Obtaining the source code	1
1.2	Installing the prerequisites	1
1.2.1	Installation on Enterprise Linux 7+	2
1.3	Installation instructions	2
2	Knowledge graph	5
2.1	Use of vocabularies	5
2.2	Notational shortcuts	5
2.2.1	Notation for typed triples	5
2.2.2	Notation for <code>rdf:List</code>	6
2.3	Datasets	7
2.4	Accounts	8
2.5	Funding	8

List of Figures

1.1	<i>Run-time references when constructed with the packages from GNU Guix.</i>	2
2.1	<i>Shorthand notation for triples with an <code>rdf:type</code> which features a hollow predicate arrow and a colored type specifier with rounded corners.</i>	6
2.2	<i>Shorthand notation for triples with a literal, which features a hollow predicate arrow and a colored rectangular type specifier.</i>	6
2.3	<i>Shorthand notation for <code>rdf:List</code> with numeric indexes, which features a hollow double-arrow. Lists have arbitrary lengths, and the numeric indexes use 1-based indexing.</i>	7
2.4	<i>The RDF pattern for a <code>djht:Dataset</code>. For a full overview of <code>djht:Dataset</code> properties, use the exploratory from the administration panel.</i>	7
2.5	<i>The RDF pattern for a <code>djht:DatasetContainer</code>. All versions of a dataset share a <code>djht:dataset_id</code> and a UUID in the container URI.</i>	8
2.6	<i>The RDF pattern for an <code>djht:Account</code>.</i>	8
2.7	<i>The RDF pattern for a <code>djht:Funding</code>.</i>	9

Chapter 1

Introduction

`djehuty` is the data repository system developed by and for 4TU.ResearchData. The name finds its inspiration in **Thoth**, the Egyptian entity that introduced the idea of writing.

1.1 Obtaining the source code

The source code can be downloaded at the **Releases**¹ page. Make sure to download the `djehuty-0.0.1.tar.gz` file.

Or, directly download the tarball using the command-line:

```
curl -LO https://github.com/4TUResearchData/djehuty/releases/\ndownload/0.0.1/djehuty-0.0.1.tar.gz
```

After obtaining the tarball, it can be unpacked using the `tar` command:

```
tar zxvf djehuty-0.0.1.tar.gz
```

1.2 Installing the prerequisites

The `djehuty` program needs Python (version 3.6 or higher) and Git to be installed. Additionally, a couple of Python packages need to be installed. The following sections describe installing the prerequisites on various GNU/Linux distributions. To put the software in the context of its environment, figure 1.1 displays the complete run-time dependencies from `djehuty` to `glibc`.

¹<https://github.com/4TUResearchData/djehuty/releases>

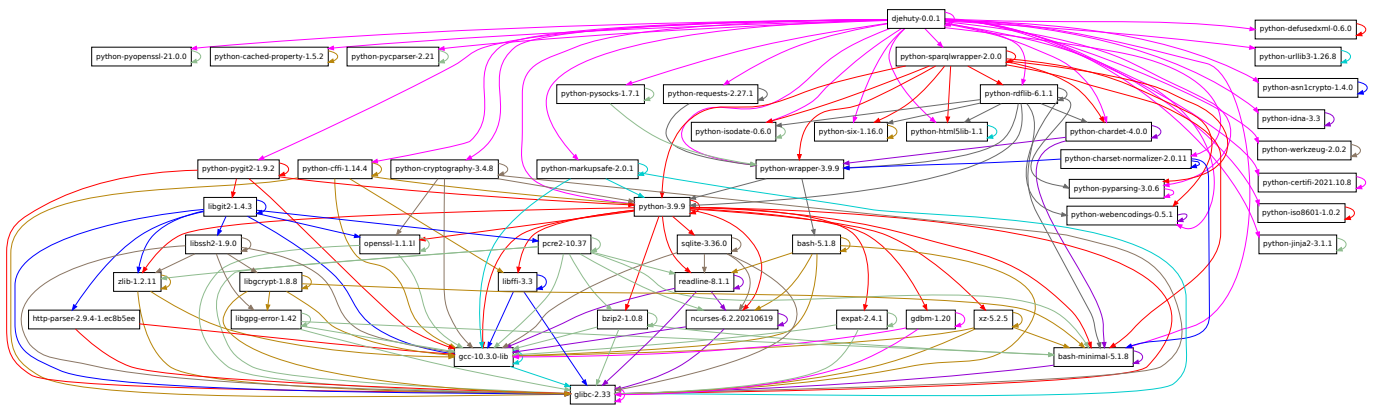


Figure 1.1: *Run-time references when constructed with the packages from GNU Guix.*

The web service of `djehuty` stores its information in a SPARQL 1.1 (“SPARQL 1.1 Overview”, 2013) endpoint. We recommend either `Blazegraph`² or `Virtuoso open-source edition`³.

1.2.1 Installation on Enterprise Linux 7+

The Python packages on Enterprise Linux version 7 or higher seem to be too far out of date. So installing the prerequisites involves two steps.

The first step involves installing system-wide packages for Python and Git.

```
yum install python39 git
```

The second step involves using Python's `venv` module to install the Python packages in a virtual environment:

```
python3.9 -m venv djehuty-env
. djehuty-env/bin/activate
cd /path/to/the/repository/checkout/root
pip install -r requirements.txt
```

1.3 Installation instructions

After obtaining the source code (see section 1.1 ‘Obtaining the source code’) and installing the required tools (see section 1.2 ‘Installing the prerequisites’), building involves running the following commands:

```
cd djehuty-0.0.1
autoreconf -vif # Only needed if the "./configure" step does not work.
./configure
make
make install
```

²<https://blazegraph.com/>

³<http://vos.openlinksw.com/owiki/wiki/VOS>

To run the `make install` command, super user privileges may be required. Specify a `--prefix` to the `configure` script to install the tools to a user-writeable location to avoid needing super user privileges.

After installation, the `djehuty` program will be available.

Chapter 2

Knowledge graph

`djehuty` processes its information using the Resource Description Framework ([Lassila, 1999](#)). This chapter describes the parts that make up the data model of `djehuty`.

2.1 Use of vocabularies

Throughout this chapter, abbreviated references to ontologies are used. Table 2.1 lists these abbreviations.

Abbreviation	Ontology URI
<code>djht</code>	https://ontologies.data.4tu.nl/djehuty/0.0.1/
<code>rdf</code>	http://www.w3.org/1999/02/22-rdf-syntax-ns#
<code>rdfs</code>	http://www.w3.org/2000/01/rdf-schema#
<code>xsd</code>	http://www.w3.org/2001/XMLSchema#

Table 2.1: Lookup table for vocabulary URIs and their abbreviations.

2.2 Notational shortcuts

In addition to abbreviating ontologies with their prefix we use another notational shortcut. To effectively communicate the structure of the RDF graph used by `djehuty` we introduce a couple of shorthand notations.

2.2.1 Notation for typed triples

When the `object` in a triple is *typed*, we introduce the shorthand to only show the type, rather than the actual value of the `object`. Figure 2.1 displays this for URIs, and figure 2.2 displays this for literals.

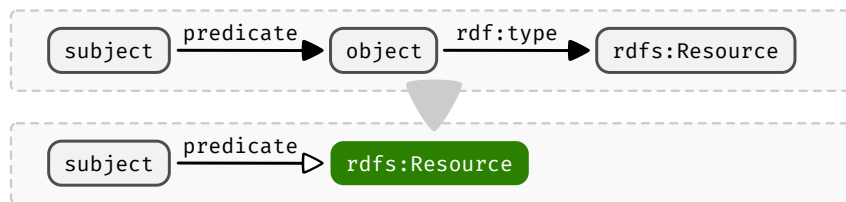


Figure 2.1: Shorthand notation for triples with an `rdf:type` which features a hollow predicate arrow and a colored type specifier with rounded corners.

Literals are depicted by rectangles (with sharp edges) in contrast to URIs which are depicted as rectangles with rounded edges.

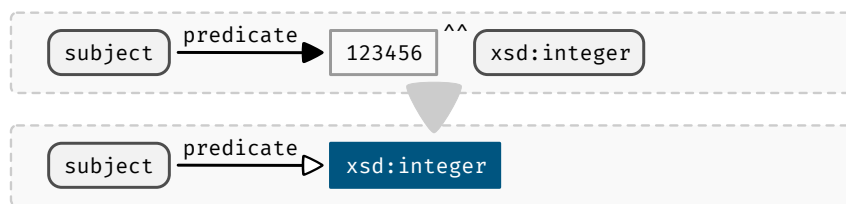


Figure 2.2: Shorthand notation for triples with a literal, which features a hollow predicate arrow and a colored rectangular type specifier.

When the subject of a triple is the shorthand type, assume the subject is not the type itself but the subject which has that type.

2.2.2 Notation for `rdf:List`

To preserve the order in which lists were formed, the data model makes use of `rdf:List` with numeric indexes. This pattern will be abbreviated in the remainder of the figures as displayed in figure 2.3.

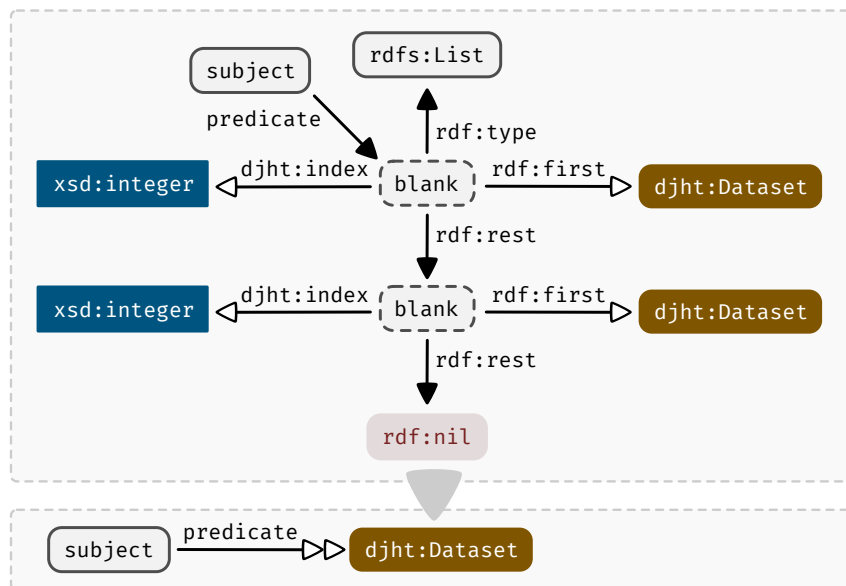


Figure 2.3: Shorthand notation for `rdf:List` with numeric indexes, which features a hollow double-arrow. Lists have arbitrary lengths, and the numeric indexes use 1-based indexing.

The hollow double-arrow depicts the use of an `rdf:List` with numeric indexes.

2.3 Datasets

Datasets play a central role in the repository system because every other type links in one way or another to it. The user submits files along with data about those bytes as a single record which we call a `djht:Dataset`. Figure 2.4 shows how the remainder of types in this chapter relate to a `djht:Dataset`.

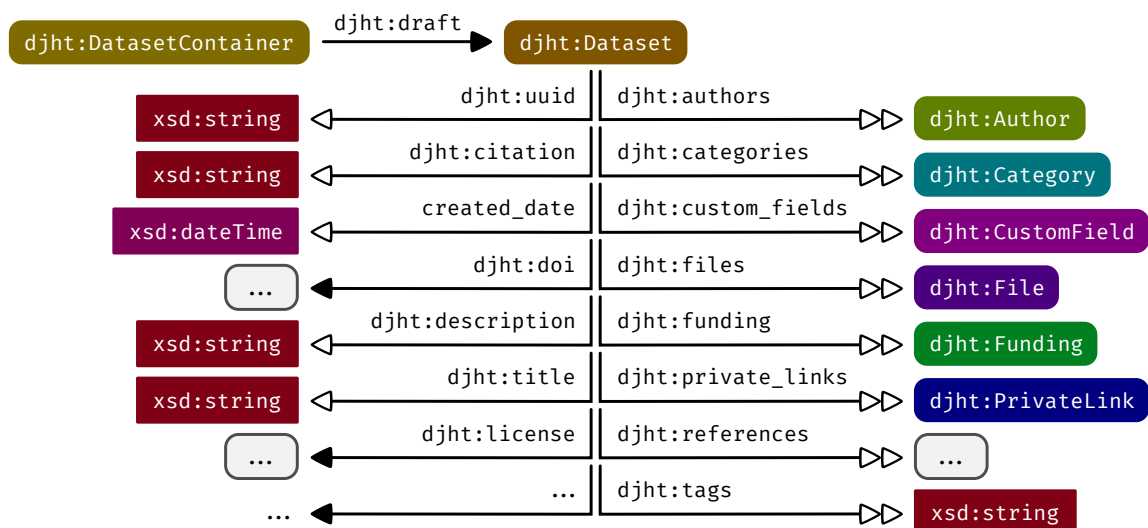


Figure 2.4: The RDF pattern for a `djht:Dataset`. For a full overview of `djht:Dataset` properties, use the exploratory from the administration panel.

Datasets are versioned records. The data and metadata between versions can differ, except all versions of a dataset share an identifier. We use `djht:DatasetContainer` to describe the version-unspecific properties of a set of versioned datasets.

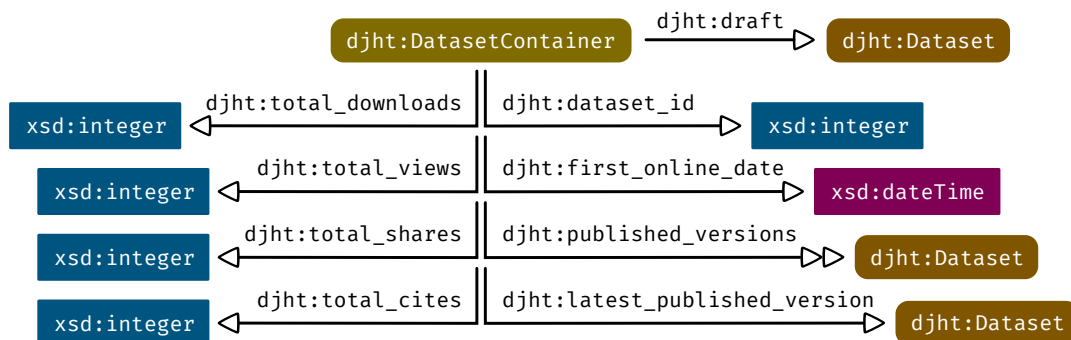


Figure 2.5: The RDF pattern for a `djht:DatasetContainer`. All versions of a dataset share a `djht:dataset_id` and a UUID in the container URI.

The data model follows a natural expression of published versions as a linked list. Figure 2.5 further reveals that the *view*, *download*, *share* and *citation* counts are stored in a version-unspecific way.

2.4 Accounts

djehuty uses an external identity provider, but stores an e-mail address, full name, and preferences for categories.

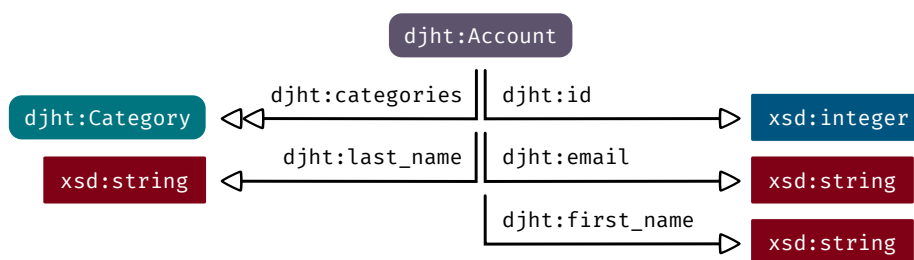


Figure 2.6: The RDF pattern for an `djht:Account`.

2.5 Funding

When the `djht:Dataset` originated out of a funded project, the funders can be listed using `djht:Funding`. Figure 2.7 displays the details for this structure.

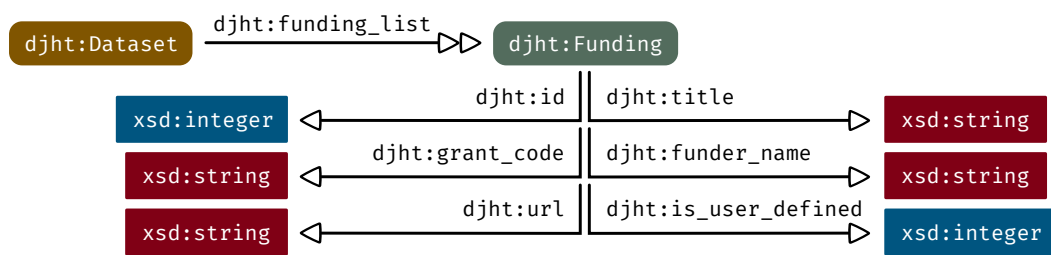


Figure 2.7: The RDF pattern for a *`djht:Funding`*.

References

Lassila, O. (1999, February). Resource description framework (RDF) model and syntax specification [W3C Recommendation]. (<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>)

SPARQL 1.1 overview [W3C Recommendation]. (2013, March). (<http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>)