

Dockerizing Linked Data

Knowledge Base Shipping to the Linked Open Data Cloud

Natanael Arndt, Markus Ackermann,
Martin Brümmer and Thomas Riechert

Semantics, Vienna 2015

September 16th 2015

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Introduction

Current Situation

- ▶ Low availability of popular SPARQL endpoints
- ▶ Setting up a local SPARQL endpoint is a complex task
- + Our approach could help to improve reproducibility of experiments with/on RDF data

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Introduction

Setting Up a SPARQL Endpoint

1. Installation of a triple store
2. Loading the data to be published into the triple store
3. Setting up a (publicly available) SPARQL endpoint
4. Providing a presentation application to support the exploration of the knowledge base
5. Ensuring de-referencability of IRIs occurring in the published knowledge base
6. Maintain the setup and ensure its availability

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Introduction

Linked Data Stack

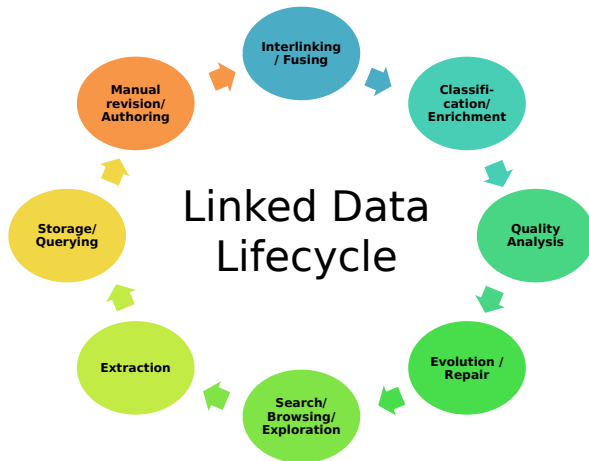


Figure: The Linked Data Stack <http://stack.linkeddata.org>

Dockerizing Linked Data

Natanael Arndt

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint

Linked Data Stack

- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Introduction

Linked Data Stack

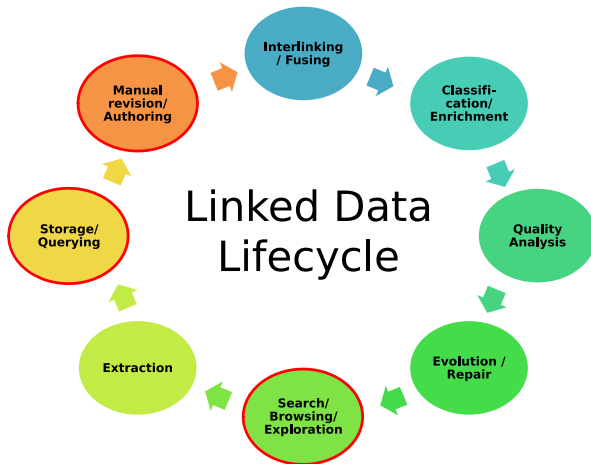


Figure: Position in the Linked Data Stack <http://stack.linkeddata.org>

Introduction

Current Situation
Setting Up a SPARQL Endpoint

Linked Data Stack

Setup a Local SPARQL Endpoint
Other Approaches
The DLD Approach

Background and Tooling

Container-Based Virtualisation
Docker
Docker Compose

Architecture

Provisioning of a Container Setup

Discussion and Conclusion

Introduction

Setup a Local SPARQL Endpoint

- ▶ The setup is a time consuming task
- ▶ Requires a certain level of technical knowledge
- ▶ Need of DevOp-competences or an administrator

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Introduction

Other Approaches

- ▶ What about a hosted Linked Data publishing infrastructure e.g. <http://dydra.com/>

Dockerizing Linked Data

Natanael Arndt

Introduction

Current Situation

Setting Up a SPARQL Endpoint

Linked Data Stack

Setup a Local SPARQL Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container Setup

Discussion and Conclusion

Introduction

Other Approaches

- ▶ What about a hosted Linked Data publishing infrastructure e.g. <http://dydra.com/>
 - ▶ Still depend on the availability of the service
 - ▶ No controlled environment
 - ▶ Privacy issues?

Dockerizing Linked Data

Natanael Arndt

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Introduction

Other Approaches

- ▶ What about a hosted Linked Data publishing infrastructure e.g. <http://dydra.com/>
 - ▶ Still depend on the availability of the service
 - ▶ No controlled environment
 - ▶ Privacy issues?
- ▶ Why aren't we just using any packaging system?

Dockerizing Linked Data

Natanael Arndt

Introduction

Current Situation

Setting Up a SPARQL Endpoint

Linked Data Stack

Setup a Local SPARQL Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container Setup

Discussion and Conclusion

Introduction

Other Approaches

- ▶ What about a hosted Linked Data publishing infrastructure e.g. <http://dydra.com/>
 - ▶ Still depend on the availability of the service
 - ▶ No controlled environment
 - ▶ Privacy issues?
- ▶ Why aren't we just using any packaging system?
 - ▶ limits in platform choice, portability

Dockerizing Linked Data

Natanael Arndt

Introduction

Current Situation

Setting Up a SPARQL Endpoint

Linked Data Stack

Setup a Local SPARQL Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container Setup

Discussion and Conclusion

Introduction

The DLD Approach

- ▶ Containerised Linked Data publication infrastructure
- ▶ Benefits regarding the maintainability and ease of setup
- ▶ Modularisation of individual components, following the micro service principle

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Background and Tooling

Container-Based Virtualisation

- ▶ Provides an isolated execution environment for applications
- ▶ Keep all necessary dependencies together to run a process
- ▶ Share the complete hardware and core Operation System (OS, in contrast to full- and para-virtualisation)
- ▶ Still each container has its own filesystem
- ▶ Can allow reproducibility of execution environments

Examples of container-based virtualisation:
FreeBSD jail, Linux Containers (LXC), Docker

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

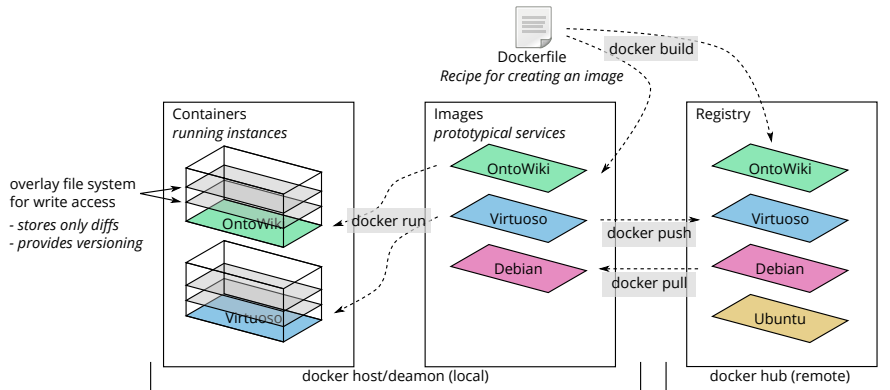
Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Background and Tooling

Docker



Docker is a container-based virtualisation system running on mayor OS: Linux, Mac OSX and Windows and on mayor cloud platforms (e.g. Amazon EC2, Microsoft Azure, Google Cloud Platform)

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation

Docker

- Docker Compose

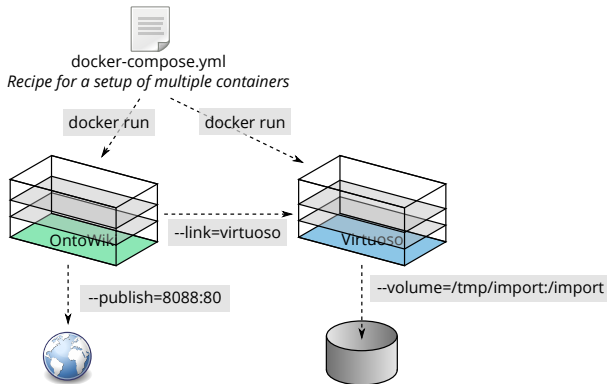
Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Background and Tooling

Docker Compose



Docker Compose (former known as *fig*), allows to combine multiple micro services to a bigger service

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker

Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Architecture

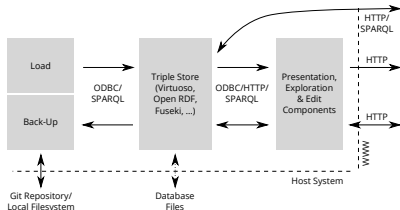


Figure: Architecture and data-flow of the containerised micro services for publishing knowledge bases

The following roles:

- ▶ Triple Storage
- ▶ Load & Back-Up
- ▶ Presentation & Publication

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Architecture

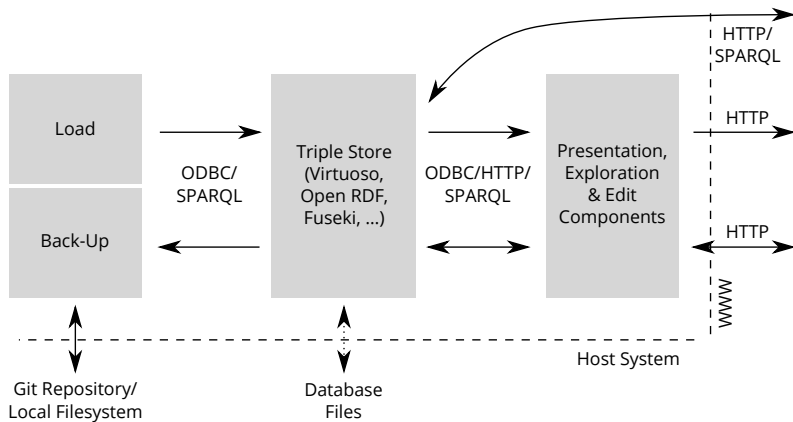


Figure: Architecture and data-flow of the containerised micro services for publishing knowledge bases

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Triple Storage

- ▶ Persistent storage is implemented using volumes
- ▶ Provides SPARQL interface
- ▶ In the current architecture there is the limitation of only one Triple Store

Load & Back-Up

- ▶ A load component is responsible for pre-loading the store
- ▶ Data is taken from a local volume or an online source
- ▶ Backup Containers dump the data to a local volume or an online repository

Presentation & Publication

- ▶ Provide exploration interfaces,
- ▶ Editing/curation interfaces or
- ▶ Domain specific views

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

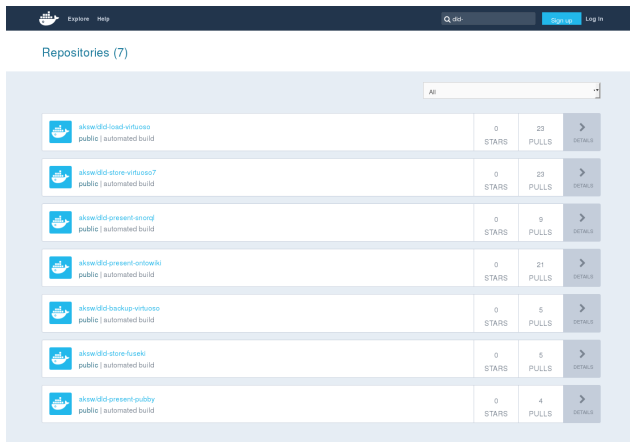
Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Architecture



The screenshot shows the Docker Hub search results for the query 'dld-'. The interface includes a top navigation bar with 'Explore' and 'Help' links, a search bar containing 'dld-', and 'Sign up' and 'Log in' buttons. Below the navigation bar, the title 'Repositories (7)' is displayed. A dropdown menu is set to 'All'. The results are presented as a table with 7 rows, each representing a repository. Each row includes a Docker Hub icon, the repository name, its visibility and build status, and columns for stars and pulls, along with a 'DETAILS' link.

Repository	Stars	Pulls	Details
akswdld-load-virtuoso public automated build	0	23	> DETAILS
akswdld-store-virtuoso7 public automated build	0	23	> DETAILS
akswdld-present-enorg public automated build	0	9	> DETAILS
akswdld-present-onowiki public automated build	0	21	> DETAILS
akswdld-backup-virtuoso public automated build	0	5	> DETAILS
akswdld-store-fuseki public automated build	0	5	> DETAILS
akswdld-present-pubby public automated build	0	4	> DETAILS

Figure: DLD containers at the docker hub, searching to "dld-"
<https://hub.docker.com/search/?q=dld-&isAutomated=0>

Dockerizing Linked Data

Natanael Arndt

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Architecture

Provisioning of a Container Setup

`datasets:`

`dbpedia-2015-endpoint-some-set:`

`graph_name: "http://dbpedia.org"`

`file: "dbp-2015-a.rdf"`

`dbpedia-2015-endpoint-some-other-set:`

`graph_name: "http://dbpedia.org"`

`location: "http://dbpedia.org/files/dbp-2015-b-
.rdf"`

`environment_global:`

`DEFAULT_GRAPH: "http://dbpedia.org"`

Listing 1: A DLD configuration file for publishing DBpedia data ...

Introduction

Current Situation

Setting Up a SPARQL
Endpoint

Linked Data Stack

Setup a Local SPARQL
Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker

Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

```
components:
  load: aksw/dld-load-virtuoso
  store:
    image: aksw/dld-store-virtuoso7
  present:
    ontowiki:
      image: aksw/dld-present-ontowiki
      ports: ["8088:80"]
```

Listing 2: ... using a setup with Virtuoso triple store, a suitable load component and OntoWiki for presentation/editing

Architecture

Provisioning of a Container Setup

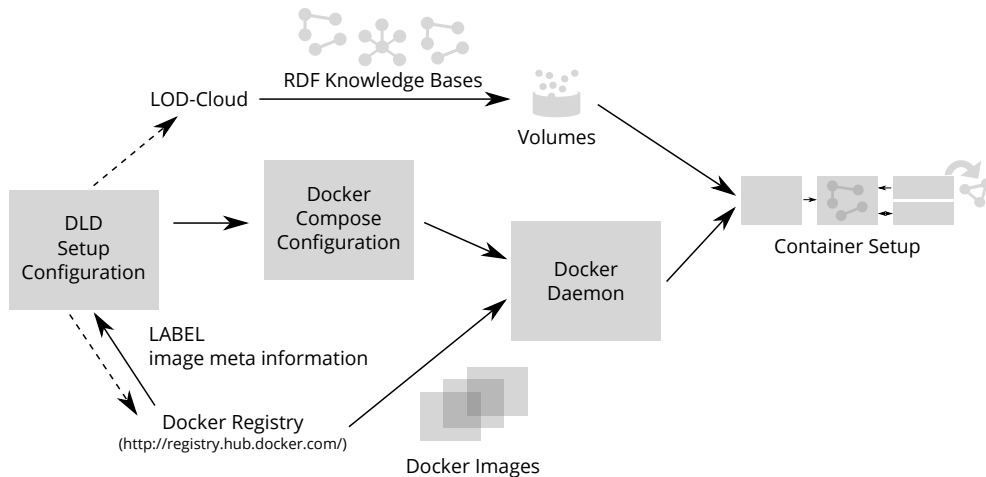


Figure: Workflow for creating a container setup

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Discussion and Conclusion

- ▶ Provide an Infrastructure for bootstrapping Linked Data publication infrastructure
- ▶ Lower the entry barrier for setting up a SPARQL endpoint
- ▶ Improved portability of Linked Data setups to transferring a `dld.yml` file

Future Work

- ▶ Embed the container and setup descriptions in DataID [1]
- ▶ Integration into the Linked Data Stack <http://stack.linkeddata.org>
- ▶ Provide more images and ready to run setups
- ▶ Improve the usability for domain experts
- ▶ Provide a UI for configuring the DLD-Setup-Tool

Introduction

Current Situation

Setting Up a SPARQL

Endpoint

Linked Data Stack

Setup a Local SPARQL

Endpoint

Other Approaches

The DLD Approach

Background and Tooling

Container-Based

Virtualisation

Docker


Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion

Discussion and Conclusion



Dockerizing Linked Data

[Introduction](#) [Requirements](#) [How To Use Bootstrap Script](#) [Available Images](#)

Popular knowledge bases that provide SPARQL endpoints for the web are usually experiencing a high number of requests, which often results in low availability of their interfaces. A common approach to counter the availability issue is to run a local mirror of the knowledge base. Running a SPARQL endpoint is currently a complex task which requires a lot of effort and technical support for domain experts who just want to use the SPARQL interface.

With our approach of containerised knowledge base shipping we are introducing a simple to setup methodology for running a local mirror of an RDF knowledge base and SPARQL endpoint with interchangeable exploration components. The flexibility of the presented approach further helps maintaining the publication infrastructure for dataset projects.

Installation

We assume you have `git`, `python` (2.7 or 3.x, with `pip`) and a `docker` daemon running. For further details please check the [requirements](#) section.

```
$ git clone https://github.com/Dockerizing/dockerizing-bootstrap
$ cd dockerizing-bootstrap
$ sudo pip install -Ur requirements.txt
```

Run

```
$ cd examples/simple/
$ ./../dld.py
$ cd wd-dld
$ docker-compose up
```

See [How To Use Bootstrap Script](#) for more advanced setup.

[▶ Run Query](#) Now your SPARQL endpoint should be available at <http://localhost:8895/sparql>.
Try to execute a simple query on your new endpoint: `select * from { ?s ?p ?o }`

Figure: Homepage with an example at: <http://dld.aks.w.org/>

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Introduction

- Current Situation
- Setting Up a SPARQL Endpoint
- Linked Data Stack
- Setup a Local SPARQL Endpoint
- Other Approaches
- The DLD Approach

Background and Tooling

- Container-Based Virtualisation
- Docker
- Docker Compose

Architecture

- Provisioning of a Container Setup

Discussion and Conclusion

Thank you for your attention!

```
$ ./dld.py
```

```
$ cd wd-dld
```

```
$ docker-compose up
```




M. Brümmer, C. Baron, I. Ermilov, M. Freudenberg, D. Kontokostas, and S. Hellmann.

DataID: Towards semantically rich metadata for complex datasets.

In Proceedings of the 10th International Conference on Semantic Systems, SEM '14, pages 84–91. ACM, 2014.

Dockerizing Linked
Data

Natanael Arndt

Introduction

Current Situation
Setting Up a SPARQL
Endpoint

Linked Data Stack
Setup a Local SPARQL
Endpoint

Other Approaches
The DLD Approach

Background and Tooling

Container-Based
Virtualisation

Docker
Docker Compose

Architecture

Provisioning of a Container
Setup

Discussion and Conclusion