

Putting Linked Data to Work: Semantic Data "Brokering" in Practice

**HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI**

**João da Silva
Stefan Negru**



OUTLINE

- ATTX Project Overview
- ATTX Semantic Broker on the ETL Landscape
 - Provenance tracking
 - Use cases
- Demo
- ATTX Technology Stack
 - CI/CD Automation
 - Deployment





SEMANTIC BROKER

USE CASE ORIENTED

UC1: Infrastructures and Publications

In house proof of concept use case.

UC2: Parallel Publication Dashboard

In cooperation with the University of Jyväskylä.

UC3: Metax

Semantic Models mapping for CSC - IT Center for Science.

UC4: MILDRED

Developing part of the University of Helsinki Research Infrastructure.



SEMANTIC BROKER

JUST ANOTHER ETL TOOL?

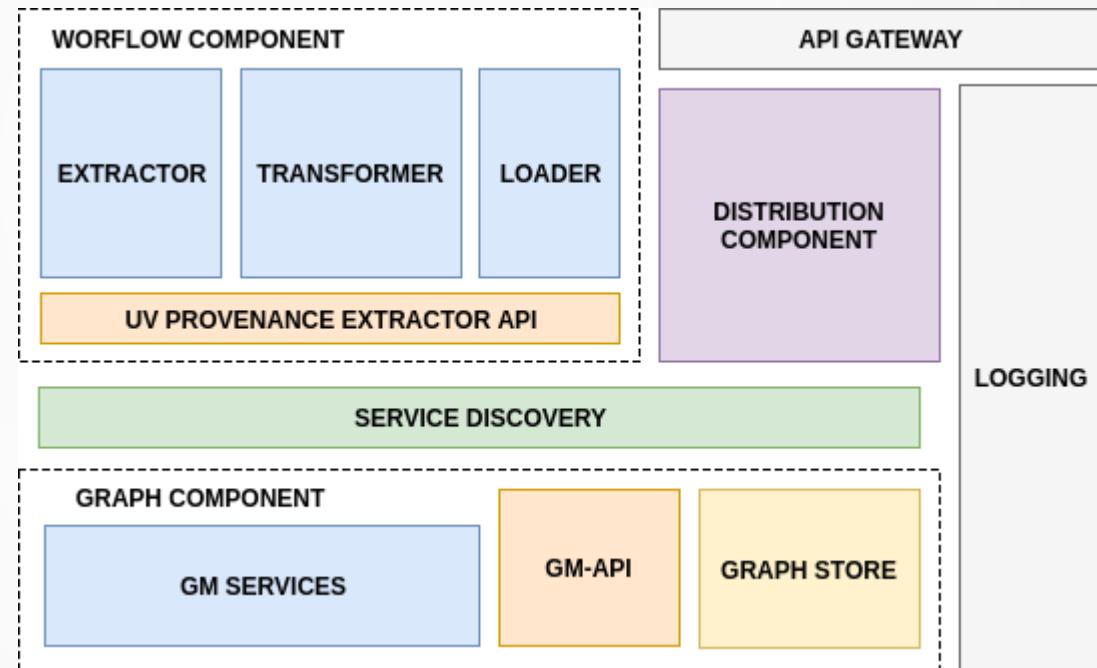
- Entity Linking (NER, reconciliation etc.)
- **Reasoning capabilities**
- Data Validation (partly including Data Quality standards)
- **Statistics and Insights about the Data**
- Data Transformations
- **Data Publishing (REST APIs, HDT files etc.)**
- **Provenance Tracking**



SEMANTIC BROKER

ATTX SEMANTIC BROKER ARCHITECTURE

- Provenance tracking
- Advance capabilities of enriching and generating new data
- Flexible graph based internal data model
- Use case oriented
- Containerised components: aiming for Micro-services Architecture





DEMO

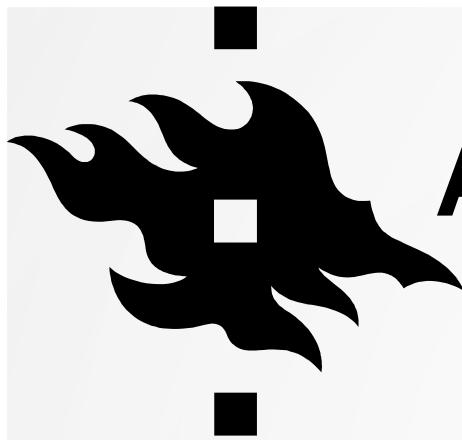
DEPLOYMENT & USAGE

- Docker Swarm as platform for ATTX components: containerise everything (and run it everywhere);
- Infrastructure-as-Code approach (yml file with service stack definition);
- Automatic deployment via YML in own PC or Cloud (OpenStack)

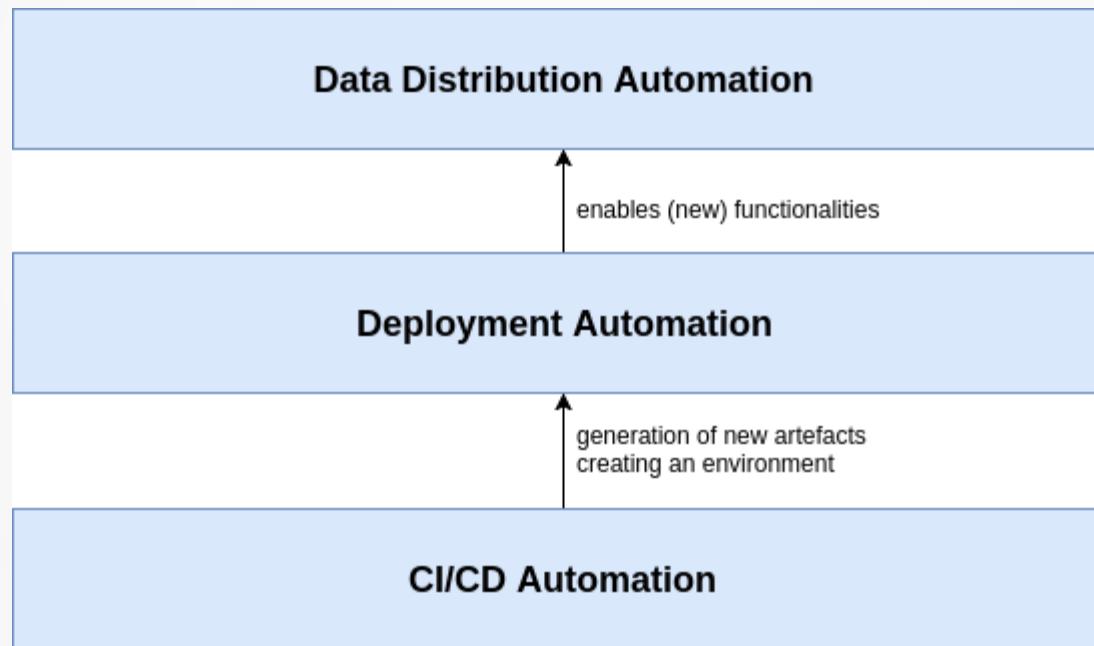


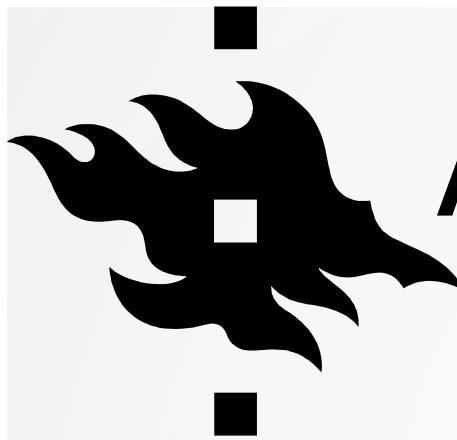
AUTOMATION FTW ...
RIGHT?

<https://www.flickr.com/photos/spacex/27294261525/>



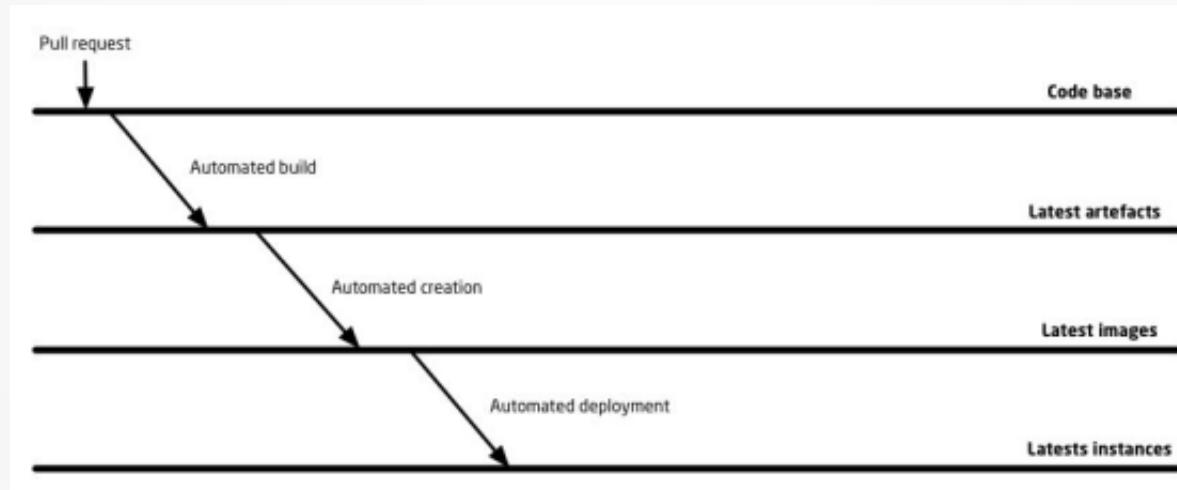
AUTOMATION LAYERS

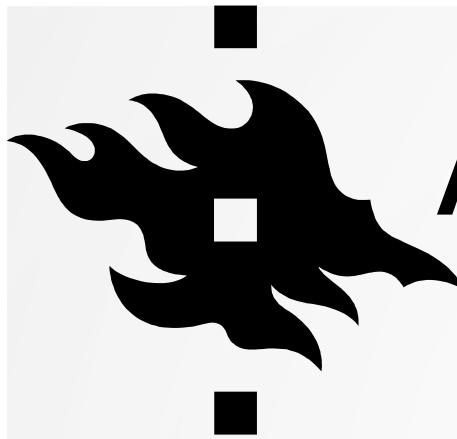




AUTOMATION LAYERS

CI/CD WORKFLOW

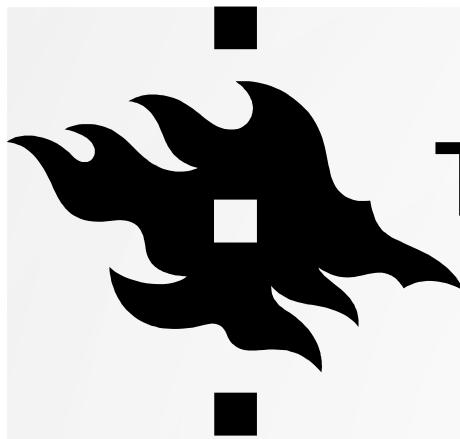




AUTOMATION LAYERS

DEPLOYMENT

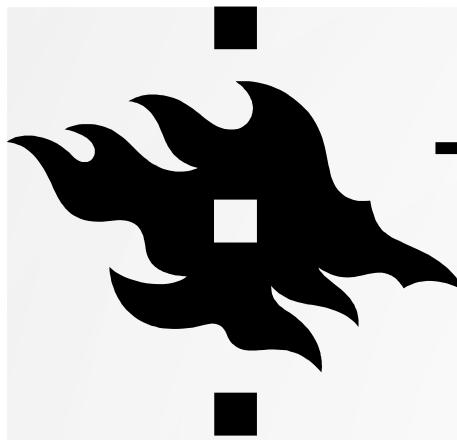
- Deploying updated components/services
- Provisioning and deploying ATTX stack on own PC or Cloud (OpenStack)
- Rolling Updates with Docker Swarm



TECHNOLOGY STACK

ATTX SEMANTIC BROKER

- ElasticSearch - <https://www.elastic.co/>
- UnifiedViews - <https://unifiedviews.eu/>
- Consul - <https://www.consul.io/>
- Jena + Fuseki - <https://jena.apache.org>
- RML - <http://rml.io/>
- SHACL - <https://www.w3.org/TR/shacl/>
- ShEx - <http://shex.io/>



TECHNOLOGY STACK

CI/CD ENVIRONMENT

- Gradle build tool - <https://gradle.org/> with {py}Gradle
- Github - <https://github.com/>
- Jenkins - <https://jenkins.io/>
- Docker - <https://www.docker.com/>
- Archiva - <https://archiva.apache.org/>



THANK YOU

"They say it could not be done,
but we did it anyway, and it (kinda) worked"

Joonas Kesäniemi - @lunch 2017

- <https://www.helsinki.fi/en/projects/attx-2016>
- <https://attx-project.github.io/>
- <https://github.com/ATTX-project/elag2017-demo>