Karlsruhe Institute of Technology

SCC HMC HELMHOLTZ METADATA COLLABORATION

A Java Library for RO-Crate

Nikola Tzochev, Andreas Pfeil (ORCID: 0000-0001-6575-1022)

Who are we? Data Exploitation Methods (DEM)

DEM is part of the IT data & computing centre of KIT

Research and development for large data collections with the goal of data exploitation

Digital Humanities

Material Science

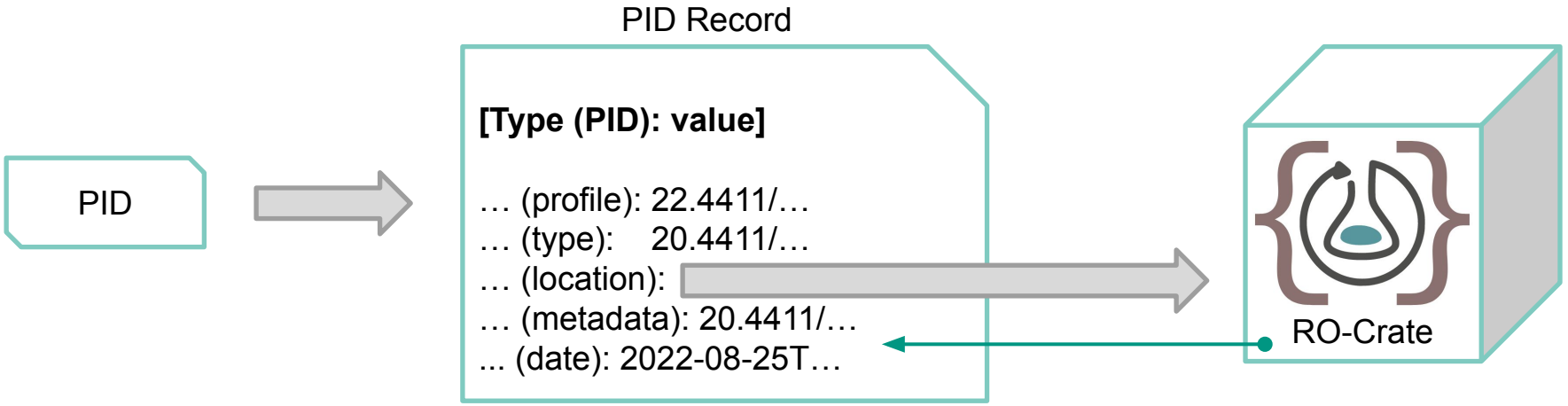
Why are we interested in RO-Crate?

- FAIR Data Commons for HMC
  - Tools and services for RDM
  - Standards and technologies supporting in the realization of FAIR
  - Focus on FAIR Digital Objects

- Interest in standardized package formats in general
  - Chaired RDA Working Group on Research Data Repository Interoperability
    - Considered RO-model and DataCrate
    - But, in the past, not found to be applicable yet

RO-Crate in a FAIR DO context

To make a FAIR DO, data must be well-described.



ro-crate-java

- Written in Java
  - Architecture different from existing implementations
  - Strategy Pattern, Interfaces
  - Partially raw JSON handling instead of direct de-serialization to a class
- Strong control through the API
  - Build valid crates and guide the user to do so
  - Builder-Pattern
- Make adding metadata easy
  - Reuse existing information (for example ORCID, ROR)

```
PersonEntity person = ORCIDProvider.getPerson("https://orcid.org/*")
OrganizationEntity organization = RORProvider.getOrganization("https://ror.org/*");
```

ro-crate-java: Example

```java
PersonEntity person = ORCIDProvider
        .getPerson("https://orcid.org/0000-002-1825-0097");

ROCrate roCrate = new ROCrate.ROCrateBuilder("name", "description")
    .addDataEntity(
        new FileEntity.FileEntityBuilder()
            .setId("survey-responses-2019.csv")
            .setSource(new File("path"))
            .setAuthor(person)
            .build()
    )
    .addContextualEntity(person)
    .build();
```

ro-crate-java: Functionality Evaluation

library name	ro-crate-py	ro-crate-ruby	ro-crate-java
version	0.6.0 (alpha)	0.4.15 (WIP)	1.0.5
Create crates	yes	yes	yes
Save/load from/to zip or folder	yes	yes	yes
Auto-generation of preview files	yes	yes	yes
Removal of entity removes references to it	no	no	yes
Warnings and errors if something is not valid or not recommended according to the specification	no	no	yes
Command line interface	yes	no	no

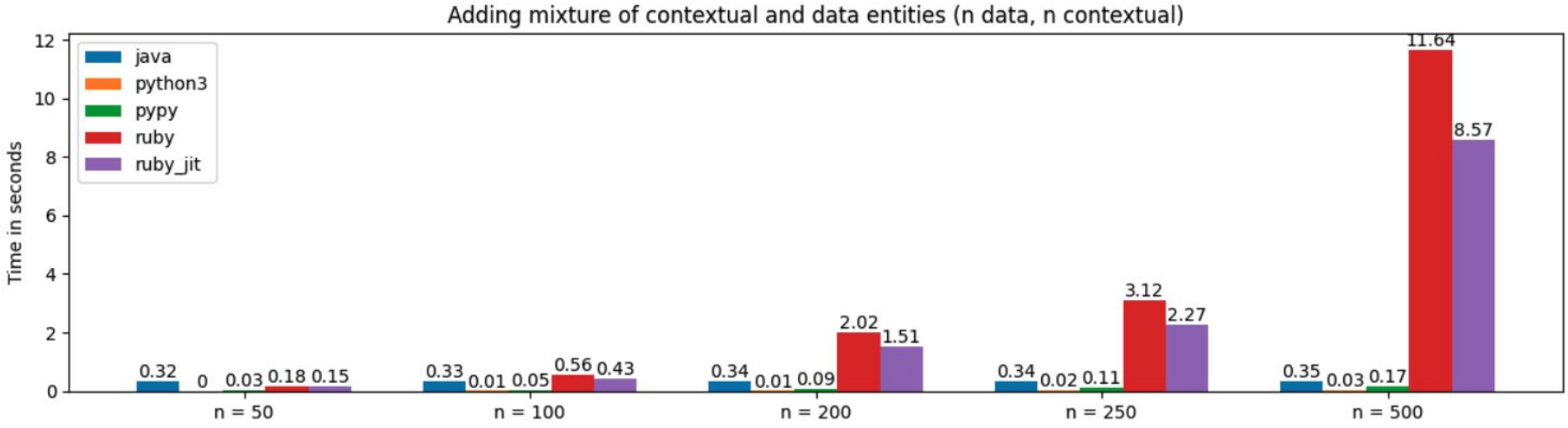
Benchmarks (excerpt)

Creating a single crate
with n contextual entities
and n data entities.

Intel i5-8265U CPU
8 GB RAM
Fedora 34

ro-crate-py v0.7.0
ro-crate-ruby v0.4.17
ro-crate-java v1.0.5
(date: 2022-08-25)

Benchmark code:
https://github.com/kit-data-manager/ro-crate-benchmarks



Conclusions

- ro-crate-java version 1.0.5 is released
- Focus on valid crates and ease of use
- Further work on it is ongoing
  - Results are adopted in various research fields of the Helmholtz Association of German research centers (e.g. in the field of energy informatics)
  - Software integrations and new developments are being evaluated

https://github.com/kit-data-manager/ro-crate-java