

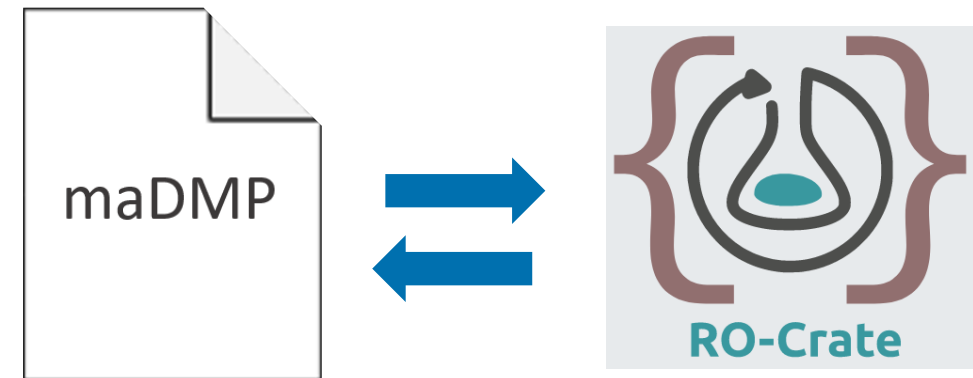
Research Object Crates and Machine-actionable Data Management Plans

Tomasz Miksa, **Maroua Jaoua**, Ghaith Arfaoui

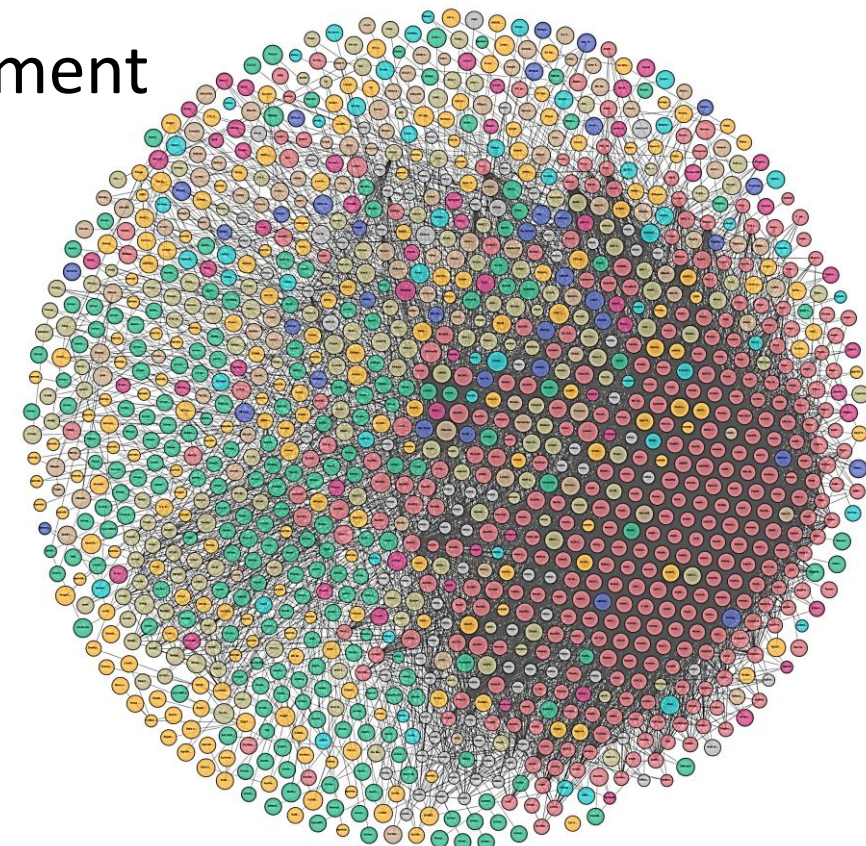
TU Wien & SBA Research, Vienna, Austria

Maroua.Jaoua@outlook.com

1. Introduction
2. Use cases
3. Mapping
4. Evaluation and Discussion
5. Conclusion



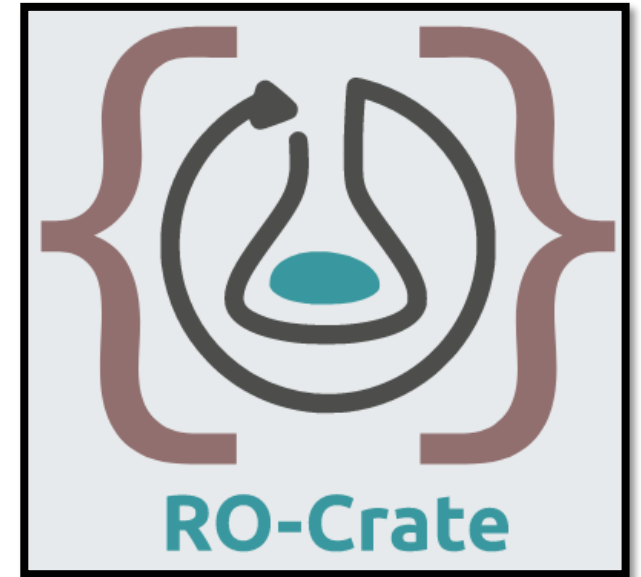
- Researchers
 - want to make their data FAIR
 - should have less work with data management
- Solutions
 - use semantic technologies
 - depend on Linked Open Data (LOD)
- Examples
 - RO-Crates
 - Machine-actionable DMPs



<https://lod-cloud.net/>

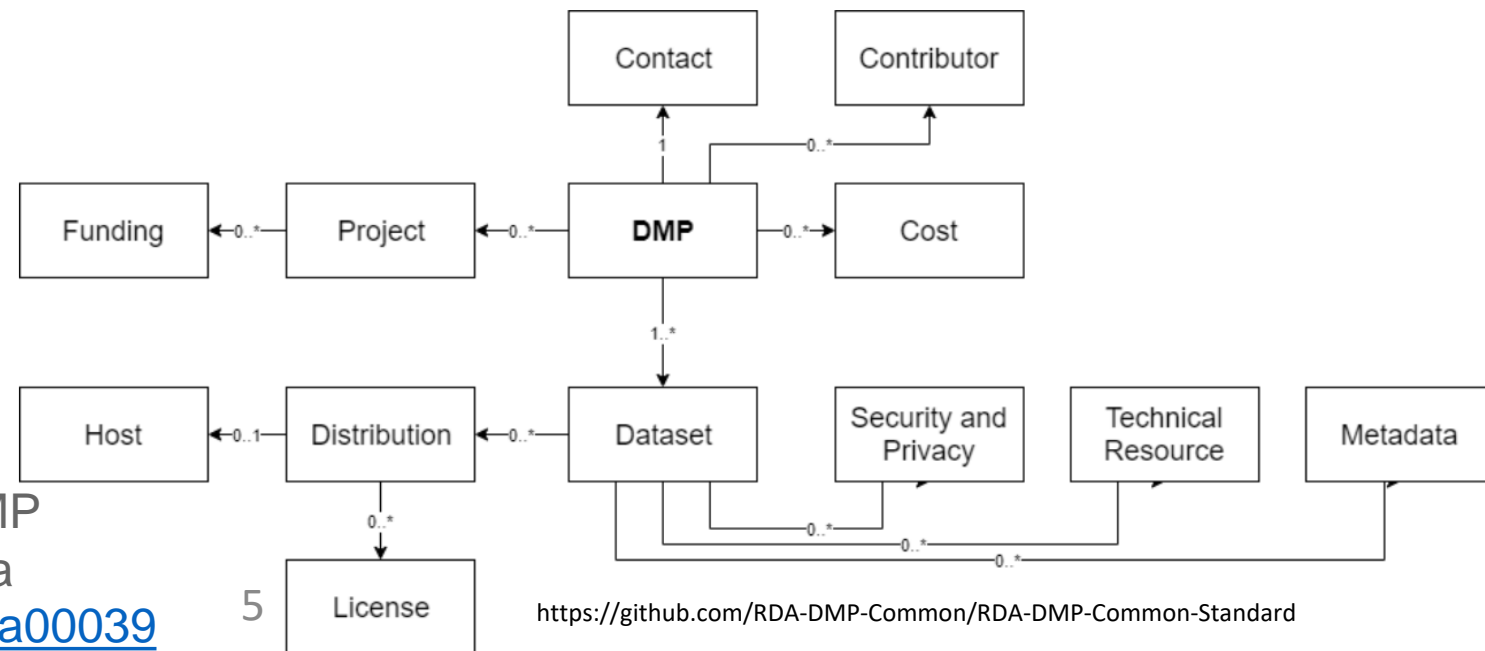
Introduction (RO-Crate)

- Organize file-based data with associated metadata
- Use linked data principles
- Researchers, digital repository managers and data stewards need them to make their data FAIR
- Simple way to
 - assert the authors (e.g. people, organizations)
 - capture complex provenance for files
- Different implementations allowing consumption and production of RO-crates



Introduction (maDMP)

- Official RDA Recommendation
- Allow exchange of DMPs between systems involved in research data management
- Represent information over the whole DMP lifecycle
- Implementations allowing the use and creation of maDMPs



Miksa, T., Walk, P., & Neish, P. (2019). RDA DMP Common Standard for Machine-actionable Data Management Plans. <https://doi.org/10.15497/rda00039>

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

- RO-crates
 - Can be created manually
 - Already exist
- ➔ RO-Crates used to
 - fill in parts of maDMPs
 - reduce workload

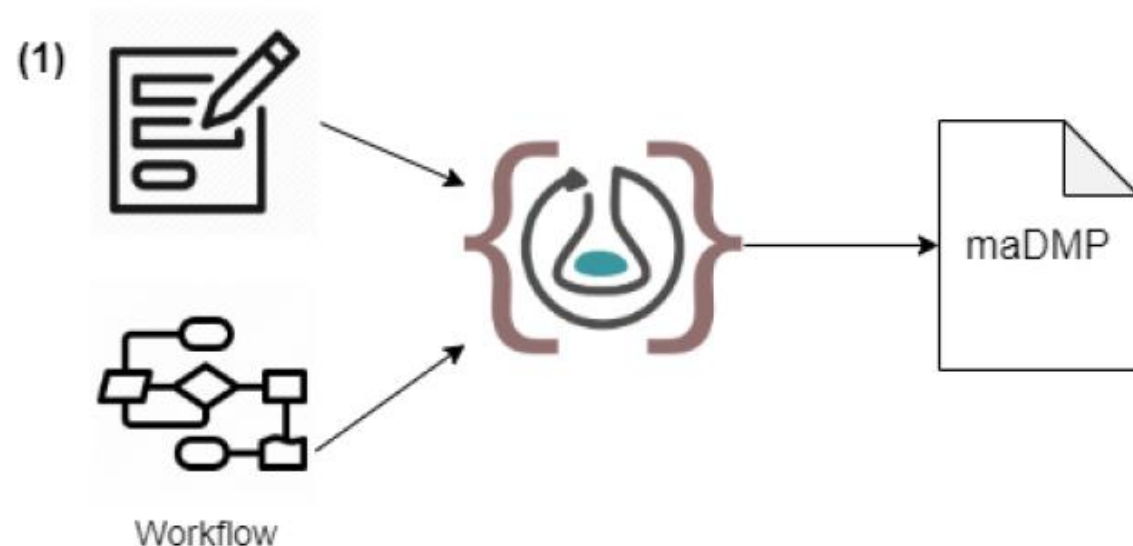


Fig. 1 Information used in RO-Crates is re-purposed for maDMPs

RO-Crates do not exist at an early stage of projects unlike maDMPs

➔ information in maDMPs can be used to facilitate creation of RO-Crates

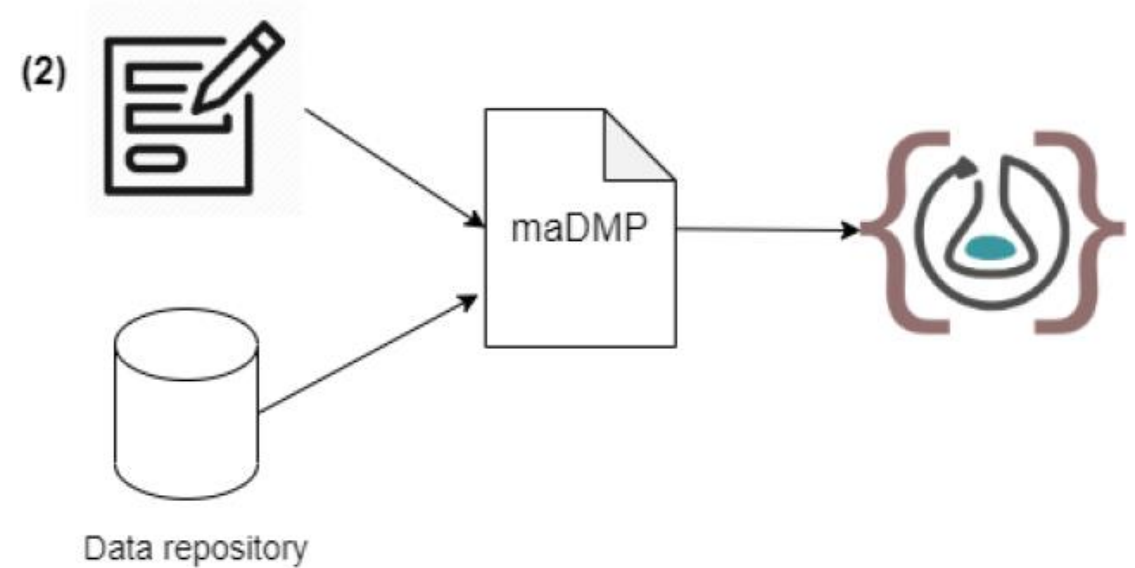
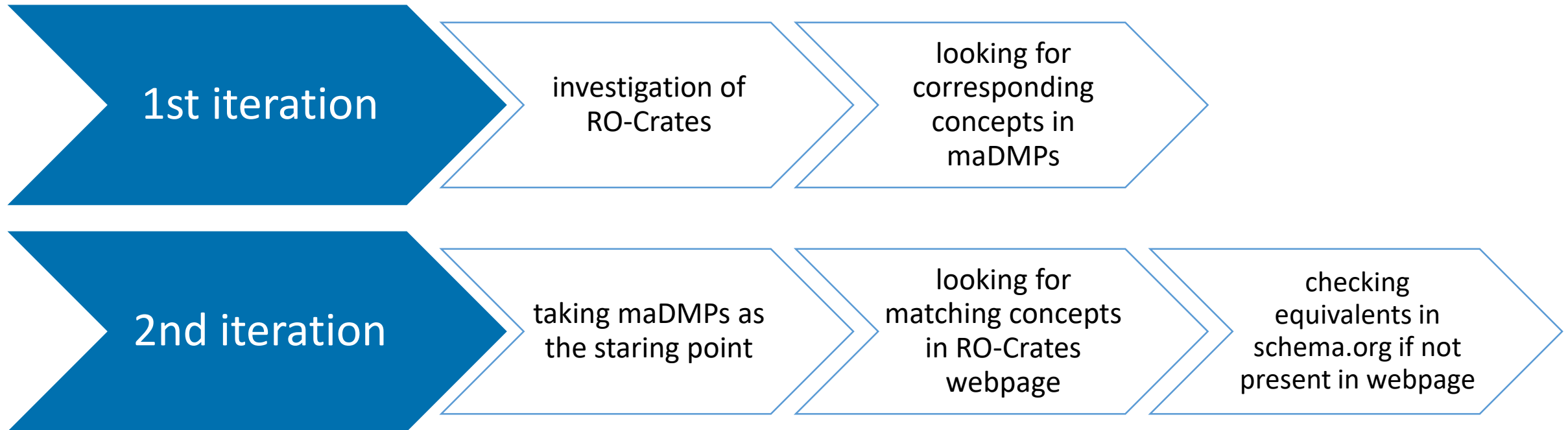


Fig. 2 maDMPs are used to generate RO-Crates

Mapping - Methodology



- 46 Mapped properties
- 7 Partially mapped properties
- 13 Unmapped RO-Crate properties
- 33 Unmapped maDMP properties
- Full mapping in repository

<https://github.com/GhaithArf/ro-crate-rda-madmp-mapper>

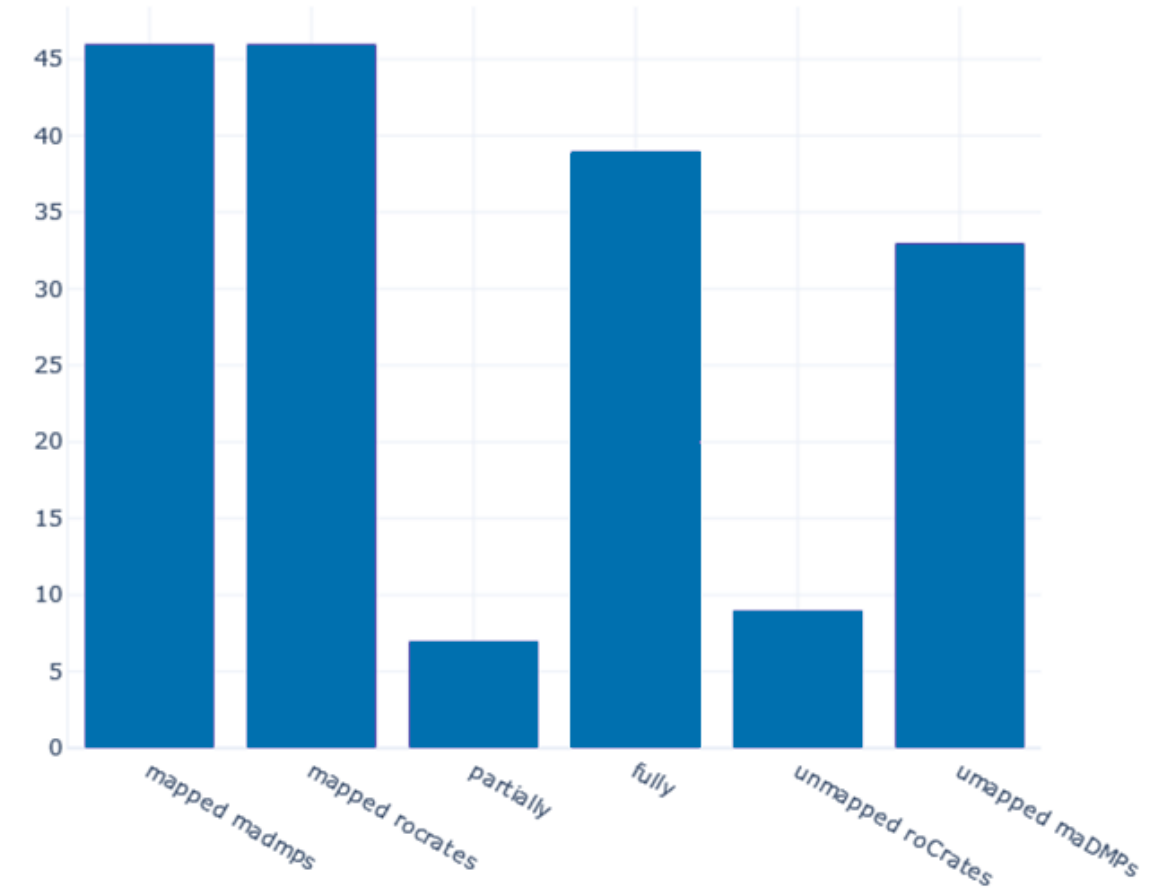


Fig. 3. Overview about the mapped and unmapped maDMP and RO-Crate properties

Mapping - Assumptions

- Significant overlap between the RO-Crates and maDMPs
- Mapped properties not always exact
- ➔ Some assumptions are introduced
 - Equivalence of the definitions of some properties
 - Equivalence of property formats

Evaluation and Discussion

- A migration tool was developed
- The tool allows generating
 - RO-Crate(s) from maDMP (1 to many)
 - maDMP from RO-Crate(s) (many to 1)
- The tool was tested with
 - 10 maDMPs from RDA DMP Common Standard repository
(<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON>)
and Data Stewardship community on Zenodo
(<https://zenodo.org/communities/tuw-dmps-ds-2020>)
 - 5 RO-Crates from Research Object Crate website
(<https://data.research.uts.edu.au/examples/ro-crate/0.2/>) and manually filled RO-Crates

```
{
  "contact": {
    "mbox": "maroua.jaoua@student.tuwien.ac.at",
    "name": "Maroua Jaoua",
    "contact_id": {
      "identifier": "https://orcid.org/0000-0001-8109-9644",
      "type": "orcid"
    }
  },
  "language": "eng",
  "ethical_issues_exist": "no",
  "dataset": [
    {
      "title": "generated data",
      "description": "Data which includes the data generated
        by running the jupyter notebook",
      "type": "document",
      "issued": "2020-03-25",
      "dataset_id": {
        "identifier": "10.5281/zenodo.3770405",
        "type": "doi"
      }
    }
  ]
}
```

Listing 1. Part of the original maDMP used to generate the RO-Crate



```
{
  {
    "@id": "https://orcid.org/0000-0001-8109-9644",
    "email": "maroua.jaoua@student.tuwien.ac.at",
    "name": "Maroua Jaoua",
    "@type": "ContactPoint"
  },
  {
    "contactPoint": {
      "@id": "https://orcid.org/0000-0001-8109-9644"
    },
    "identifier": "10.5281/zenodo.3770405",
    "description": "Data which includes the data generated by
      running the jupyter notebook",
    "hasPart": [
      {
        "@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/igo/"
      }
    ],
    "datePublished": "2020-03-25",
    "name": "generated data",
    "Language": "eng",
    "@type": "Dataset",
    "@id": "./"
  }
}
```

Listing 2. Example of a generated RO-Crate based on maDMP

- RO-Crates and maDMPs can be mapped to enable exchange of information between them
- A significant number of properties can be mapped directly
- Such automated mapping can help exchange information between RO-Crates and maDMPs efficiently
- The future work
 - further automation of RO-Crate and maDMP creation
 - the evaluation of further examples
 - extend the conversion tool with a support for maDMPs serialized using ontologies

1. Miksa, T., Walk, P., Neish, P.: **RDA DMP Common Standard for Machineactionable Data Management Plans** (2019). <https://doi.org/10.15497/rda00039>
2. O Carragain, E., Goble, C., Sefton, P., Soiland-Reyes, S.: **A lightweight approach to research object data packaging. Bioinformatics Open Source Conference (BOSC2019)** (Jun 2019). <https://doi.org/10.5281/zenodo.3250687>
3. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., HillmanJackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D.: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update**. Nucleic Acids Research 46(W1), W537–W544 (05 2018). <https://doi.org/10.1093/nar/gky379>
4. Miksa, T., Neish, P., Walk, P., Rauber, A.: **Defining requirements for machine-actionable data management plans**. In: McGovern, N., Whiteside, A. (eds.) Proceedings of the 15th International Conference on Digital Preservation, iPRES 2018, Boston, MA, USA, September 24-28, 2018 (2018), <https://hdl.handle.net/11353/10.923628>
5. Miksa, T., Simms, S., Mitchen, D., Jones, S.: **Ten principles for machine-actionable data management plans**. PLOS Computational Biology 15(3), 1–15 (03 2019). <https://doi.org/10.1371/journal.pcbi.1006750>,