

RDA DMP Common Standard for Machine-actionable Data Management Plans



DOI: [10.15497/rda00039](https://doi.org/10.15497/rda00039)

Authors: Tomasz Miksa, Paul Walk, Peter Neish

RDA Working Group: DMP Common Standards

Published: 16th September 2020

Abstract: Data Management Plans are free-form text documents describing the data that is used and produced during the course of research activities. They specify where the data will be archived, which licenses and constraints apply, and to whom credit should be given, etc. The workload and bureaucracy often associated with traditional DMPs can be reduced when they become machine-actionable.

The DMP Common Standards WG has developed an application profile that allows to express information from traditional DMPs in a machine-actionable way. It allows for automatic exchange, integration, and validation of information provided in DMPs.

Thus, it facilitates exchange of information between systems acting on behalf of stakeholders involved in the research life cycle, such as, researchers, funders, repository managers, ICT providers, librarians, etc.

This document summarises the application profile. The full specification can be found: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>.

Keywords: DMPs, maDMPs, machine-actionable, application profile

Language: English

License: [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](https://creativecommons.org/licenses/by/4.0/)

RDA webpage: <https://www.rd-alliance.org/group/dmp-common-standards-wg/outcomes/rda-dmp-common-standard-machine-actionable-data-management>

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

RDA DMP Common Standard for Machine-actionable Data Management Plans

Recommendation of the RDA DMP Common Standards WG
Tomasz Miksa, Paul Walk, Peter Neish

Purpose

This application profile is meant for exchange of machine-actionable DMPs between systems. It is independent of any internal data organisation used by these systems. The application profile does not prescribe how information must be presented to the end user and does not enforce any specific logic on how this information must be collected or used. The application profile is an information carrier and the full machine-actionability can only be achieved when systems using the application profile implement appropriate logic.

This application profile is intended to cover a wide range of use cases and does not set any business (e.g. funder specific) requirements. It represents information over the whole DMP lifecycle, that is, it can express planned actions, as well as actions already performed.

The application profile is NOT intended to be a prescriptive template or a questionnaire, but to provide a re-usable way of representing machine-actionable information on themes covered by DMPs. It also does not require public exposure of DMPs.

Overview

Figure 1 presents concepts used within the application profile. Each concept is further broken down into specific fields (not depicted). The full application profile specification can be found [online](#). Below we outline main concepts used within the application profile that are depicted in Figure 1.

DMP - Provides high level information about the DMP, e.g. its title, modification date, etc. It is the root of this application profile.

Project - Describes the project associated with the DMP, if applicable. It can be used to describe any type of project: that is, not only funded projects, but also internal projects, PhD theses, etc.

Funding - For specifying details on funded projects, e.g. NSF or EC funded projects.

Contact - Specifies the party which can provide information on the DMP.

Contributor - For listing all parties involved in the process of data management described by this DMP, and those parties involved in the creation and management of the DMP itself.

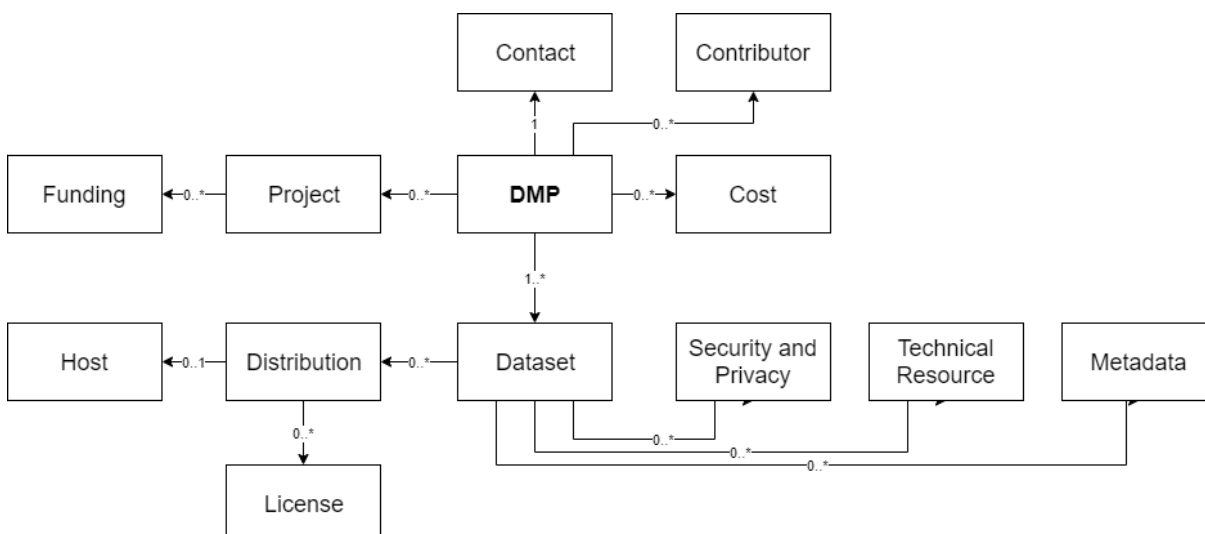


Figure 1: Overview of the application profile for the machine-actionable DMPs.

Cost - Provides a list of costs related to data management.

Dataset - This follows the definition of Dataset in the W3C DCAT specification. Dataset can be understood as a logical entity depicting data, e.g. raw data. It provides high level information about the data.

Distribution - The term "distribution" used here is as defined by the very widely used W3C DCAT metadata application profile. It is used to mean a particular instance of a dataset that has been, or is intended to be, made available in some fashion. It is important to separate the logical notion of a "dataset" from its distributions, of which there may be several, especially to attach more specific metadata properties such as "size" and "license". The lifecycle of the DMP has no particular bearing on this, and a "distribution" may be defined even if the DMP is never actually realised.

License - Used to indicate the license under which data will be made available. It also allows for modelling embargoes.

Host - Provides information on the system where data is stored.

Security and Privacy - Used to indicate any specific requirements related to security and privacy of a specific dataset.

Technical Resource - For specifying equipment needed/used to create or process the data.

Metadata - Provides a pointer to a metadata application profile used to describe the data. It does not contain any actual metadata relating to the dataset.

Methodology

The group performed an open stakeholder consultation to define the scope of information covered by machine-actionable DMPs [1]. We identified typical processes that use information from maDMPs and developed prototypes that demonstrate how typical data management tasks can be automated [2] [3]. All these actions helped us in defining this

application profile and also led to formulation of 10 principles for implementing machine-actionable DMPs [4]. Using this application profile helps in making data FAIR [5].

Contact

Participate by adopting the application profile in your setting. Sign up to the [RDA DMP Common Standards working group](#) to stay up to date. Contact chairs in case of any questions.

Important links

Full specification

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

FAQ

<https://git.io/JeX81>

JSON Examples

<https://git.io/JeX8y>

Slides, prototypes, etc.

<https://git.io/JeX8S>

References

[1] Tomasz Miksa, Peter Neish, Paul Walk, Andreas Rauber: *Defining requirements for machine-actionable Data Management Plans*. iPres 2018, Boston, USA, [DOI](#)

[2] Tomasz Miksa, João Cardoso, José Luis Borbinha: *Framing the scope of the common data model for machine-actionable Data Management Plans*. IEEE BigData 2018, [DOI](#)

[3] Asztrik Bakos, Tomasz Miksa, Andreas Rauber: *Research Data Preservation Using Process Engines and Machine-Actionable Data Management Plans*. TPD 2018, [DOI](#)

[4] Tomasz Miksa, Stephanie Simms, Daniel Mietchen, Sarah Jones: *Ten principles for machine-actionable data management plans*. PLOS Computational Biology (2019), [DOI](#)

[5] Sarah Jones, Robert Pergl, Rob Hooft, Tomasz Miksa, Robert Samors, Judit Ungvari, Rowena I. Davis, and Tina Lee. *Data Management Planning: How Requirements and Solutions are Beginning to Converge*. Data Intelligence (2019), [DOI](#)