
Data Management Tool

Lecture: Data Management in Quantitative Biology

Names: Sven Fillinger, Simon Heumos, Judith Neukamm

June 11, 2015

1 INTRODUCTION AND MOTIVATION

The increasing amount of data and the participation of several institutions in a project makes it important to document well how to handle essential aspects like the storage, analysis and integration of data during the project. This can be realized with a Data Management Plan [3]. Moreover, this plan ensures before the data collection starts that data are in correct format, well-organized and better annotated [4]. The documentation of the different steps throughout the data's life cycle helps other to understand and use the data in the future.

Furthermore, the data management plan also makes the data available to other researchers upon project completion, which can impinge positive on the whole work, concerning discovery and relevance [4].

There exists no standardized guidance how to create a data management plan however the DAMA Data Management Body of Knowledge [2] provides a good orientation of essential aspects which should be part of the plan [3].

During the scope of the project, it was our task to develop a Data Management Planning Tool. The tool should be able to create automatically a DMP based on an experimental design given as a .tsv file. This file was generated by *QWizard* [1], a portlet to input experimental data. The tool also offers users the possibility to add project information which are not included in the .tsv file.

The following chapters give an overview about the tool *CMPcreator* which was implemented during the scope of this project. The last chapters compare our tool with the already existing tool *DMPTool* developed by the University of California (wie zitieren?) followed by an outlook how our tool can be extended.

Settings for Data Management Plan

General Information

Provide some general information for your data management plan

Project Name: FIFA against the world.

Institute / Organization: FIFA

Person in Charge: Sepp Platter

Street: Im Strafraum15

General Project Description: How to get busted for payoff.

Experiment Design Upload from QWizard.

Browse... No file selected. Upload File

Uploaded File: TEACHING_DMQB_PROJECT_QUQLX.tsv

ZIP-code: 72070 City: Tubingen

Country: Germany

General Roles & Responsibilities

Figure 2.1: *General Information* Slide of DMPcreator. The progress bar is placed on the top. Fields that are fillable by the user can be seen below. Note, a special upload field for the .TSV file from Q-Wizard is visible on the left bottom.

2 RESULT: DMPCREATOR

The result of our implementation efforts is the user friendly webinterface DMPcreator. The tool is structured into five slides. Every slide...

The first slide, the *General Information Slide 2.1*, provides fields letting the user enter general information about his project, for example the project name, the person in charge of the project and contact data. Furthermore, the user can upload his .TSV file created by Q-Wizard. One important topic that needs to be covered when creating a data management plan is *Roles & Responsibilities 2.2*. This second slide allows the user to assign roles to persons. The chosen values are added two a responsibilities list. Having specified who is responsible for which data, the user still has to decide, how the data is stored. Here comes the third slide *ContentManagement2.3* in handy. Here the user can assign file types to an associated description. This so built content is then added to a content table.

General Roles & Responsibilities Content Management Storage/Backup Dissemination

Settings for Data Management Plan

Roles & Responsibilities

Provide some general information for your data management plan

Select your role type. Asshole ▼

Person In Charge.

Add unfamiliar role. Add Responsibility

Already chosen responsibilities.

Role_Type	Person_In_Charge
Asshole	Sepp Platter.

Delete Responsibility

About Roles and Responsibilities
Scientists being aware of their roles & responsibilities maintain an efficient and productive working environment for everyone.

General Content Management

Figure 2.2: *Roles & Responsibilities* Slide of DMPcreator. The progress bar is placed on the top. Fields that are fillable by the user can be seen below.

General Roles & Responsibilities Content Management Storage/Backup Dissemination

Settings for Data Management Plan

Content Management

Please specify here which data types including content will be occurring during the project.

Select your data type. CASH ▼

Description

Add unfamiliar data type. Add Content

Already chosen contents.

Datatype	Description
CASH	Sepp knows what's good for him.

Remove Content

About Content Management
Having a clear overview over current work progresses and processes is the role of content management.

Roles & Responsibilities Storage/Backup

Figure 2.3: *Content Management* Slide of DMPcreator. The progress bar is placed on the top. Fields that are fillable by the user can be seen below.

General Roles & Responsibilities Content Management **Storage/Backup** Dissemination

Settings for Data Management Plan

Storage and Backup

[This section covers the topic of data storage/backup and archive.](#)

Storage Location
Sand

Select your backup solution.
RAID 10

Select your archive solution.
TAPE

Approximate disk space in GB needed for one PEPTIDES experiment:
200

Approximate disk space in GB needed for one DNA experiment:
200

Total space in GB needed: 16000.0
Required space for storage/backup chosen RAID solution in GB: 32000.0

About Storage and Backup
The determination of rules for storage & backup contributes to a complete data management plan.

Content Management Dissemination

Figure 2.4: *Storage & Backup* Slide of DMPcreator. The progress bar is placed on the top. Fields that are fillable by the user can be seen below.

General Roles & Responsibilities Content Management Storage/Backup **Dissemination**

Settings for Data Management Plan

Dissemination Methods

[Provide some information for your data management plan concerning the sharing and access rules of your data.](#)

Select your dissemination method.
Web

Description

Add unfamiliar method.

Add Method

Generate Report

Already chosen methods.

Method	Description
Web	Trollolo

Delete Method(s)

Download Report

Storage/Backup Dissemination

Figure 2.5: *Dissemination* Slide of DMPcreator. The progress bar is placed on the top. Fields that are fillable by the user can be seen below.

REFERENCES

- [1] Andreas Friedrich, Erhan Kenar, Oliver Kohlbacher, and Sven Nahnsen. Intuitive web-based experimental design for high-throughput biomedical data. *BioMed Res Int*, 2015:958302, 2015.
- [2] DAMA International. *The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK*. Technics Publications, LLC, USA, 2009.
- [3] Sven Nahnsen. 'lecture notes in data management in quantitative biology', April 2015.
- [4] MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Data Management. <http://libraries.mit.edu/data-management/plan/why/>. [Online; accessed 11-June-2015].