Bagman – A Tool that Supports Researchers Archiving Their Data

Claus Zinn

CLARIN 2021 Virtual Conference
Motivation

• RDM essential for good scientific practise
  • theory builds on data
  • theory verification also means data inspection
• RDM often feels like house-hold chores
  • inventory of all data
  • cleaning-up data, structuring data
  • documentation of procedures and scripts
  • metadata provision
  • bundling and archiving
• Tool required to support RDM
Background

1. Data Packaging
   - thinking of common / bad practise…
   - Docuteam Packer (software), BagIt (specification)

2. Metadata Assignment
   - CMDI as metadata infrastructure framework
   - Complexity makes it hard to use to its full potential

3. Archiving Procedure
   - where should my data best go, e.g., to maximise visibility
Bagman

- Web-based software
- Supports Researchers
- Supports Archive Managers

Bagman helps you package your research data and transfer it safely to your archive manager.

Your data will be shared with the archive manager only. For more details, see the FAQ.
Bagman

- Metadata collection via forms
- Different categories
- Predefined profiles
- Fixed set of licences

Pack your data

Use the seven tabs below to describe your research data and to pack it into an archive.

Information about your project

Please provide the following information about your project. If your project affiliation is "internal" or "external", use the project description to make it more concrete. Press the Submit button to register your information with the system.

Project Acronym
PARROT

Project Name
Second Language Acquisition in Parrots

Project Affiliation
SFB 833
Bagman – Demonstration

- [https://weblicht.sfs.uni-tuebingen.de/bagman/](https://weblicht.sfs.uni-tuebingen.de/bagman/)
- fully functional prototype version
  - only TALAR archive is connected
  - other archives, other profiles to be included
    - logins for archive managers
    - CMDI generation (and additional forms) for new archives and their requirements
- initial feedback encouraging
  - e.g., now browser session can be saved and later restored to continue with RDM
Bagman – Discussion

- design walks fine line between researchers and archive managers
  - for one side, the more metadata, the better
  - for the other side, get it done quickly…
  - some data mandatory, some other data can be added voluntarily
- no direct editing of CMDI
  - entire CMDI is generated from form-based input
  - and from the data being archived
- no ideological warfare
  - e.g., granularity of data being archived, or data formats being accepted
  - Bagman as the middle man, also as a means to establish communication