



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](#)

[Pack your data](#)

[Archive HQ](#)

[Help](#)

[Registered Archives](#)



Bagman

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](#).

[Pack your data](#)

[Archive HQ](#)

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.

[Project](#)

[Organisation](#)

[Researcher\(s\)](#)

[Profile](#)

[Licence](#)

[Your Data](#)

[Submission](#)

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

Uploaded research data

Note. Files with red icons use file formats unsuitable for archiving

Filter with:

- data
 - AlexanderVonHumboldt
 - observations
 - Perroquet_Dutvowel.xls
 - Perroquet_AECons.xls
 - Perroquet_AEVowe.xls
 - Perroquet_Dutconso.xls
 - documentation
 - Coding_Parrots.pdf
 - vonHumboldt_NouveauContentient1814.pdf
 - Coding_Parrots.docx
 - vonHumboldt_Essai_Perroquet.pdf
 - vonHumboldt_NouveauContinent1814.docx
 - vonHumboldt_Essai_Perroquet.docx

Bag Info

Label	Value
Source-Organization	Eberhard Karls Universität Tübingen
Contact-Name	Alexander von Humboldt
Contact-Phone	+49 (0) 7071-29 73968
Contact-Email	avh@uni-tuebingen.de
Description	Second Language Acquisition in Parrots
Bagging-Date	2021-04-27
BagIt-Version:	1.0
Tag-File-Character-Encoding:	UTF-8
Bag-Count:	10
Bag-Size:	18.9 MB

I accept the terms and conditions (link follows...).

Bag Entries

File	Size	Mimetype	SHA256
AlexanderVonHumboldt/observations/Perroquet_AEVowe.xls	4835840	application/vnd.ms-excel	c4c9aa805fda4eea2dc1aed77638520427290f6bdd8d57d2ab3c2a99ce7c7c9a
AlexanderVonHumboldt/documentation/Coding_Parrots.pdf	55561	application/pdf	446c8f51286e25015270ff054beeafa68ab9fb8be537acb96f0f30c29dec0819
AlexanderVonHumboldt/documentation/vonHumboldt_NouveauContentient1814.pdf	99506	application/pdf	1c991c1f2599ebcc5ba82927b7382e64f4027b0f91cd8a450c5fc609a5c6e3c2
AlexanderVonHumboldt/documentation/Coding_Parrots.docx	14822	application/vnd.openxmlformats-officedocument.wordprocessingml.document	9dff7c8802debb5315301d46513a5dd6e207c38c14e09ae794f2f4fb0fa85518
AlexanderVonHumboldt/documentation/vonHumboldt_Essai_Perroquet.pdf	133324	application/pdf	7007172cf4807c11d17ae7a6bb204850d69ebdfb50094a2f7574724bc123f7e4



Bagman – Demonstration

- <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