

Datenmanagement als Voraussetzung für Open Access – Ein Fallbeispiel

Dr. Ana Sesartic

Digitaler Datenerhalt, ETH-Bibliothek, ETH Zürich

ana.sesartic@library.ethz.ch www.library.ethz.ch/Digitaler-Datenerhalt



Open-Access-Tage 2015 Zürich

1 Einleitung

Die seit 1988 bestehende Fachgruppe **Systemökologie** an der **ETH Zürich**, hat unter der Leitung von Prof. Andreas Fischlin seit ihrer Gründung aktiv Forschungsdatenmanagement (FDM) betrieben.

Wir zeigen, was sich in diesen Jahrzehnten des FDM bewährt hat, welche **Erfahrungen** gemacht wurden und welche Probleme es zu lösen gilt.

2 Archivkonzept

Forschungsschwerpunkte der Fachgruppe Systemökologie waren u.a. der Einfluss des Klimawandels auf Ökosysteme, Systemtheorie und Modellierung.

Dabei entstehende **Datentypen** waren wissenschaftliche Publikationen, Grafiken, Datenreihen aus Beobachtungen und Modellsimulationen sowie in der Gruppe entwickelte Software.

```
TITLE.....RAMSES 2.2b6 (haikal)
VERSION.....2.2b6
LIST_OF_PARTS.....:00_Index (/pub/mac/RAMSES/)
:00_README.RAMSES
:00Manual_2.2b.sea.hqx
:RAMSES 2.2b6.sea
:RAMSES_2.2b6.sea.hqx
AUTHORS.....RAMSES developers
DESCRIPTION.....RAMSES 2.2b6 as found on URL
"ftp://ftp.isg.umw.ethz.ch/pub/mac/RAMSES/"
("haikal.ethz.ch), before it was removed the 5th
of July 1996.
HANDLING.....BinHex 4.0 and MacBinaryII-files of Compact Pro self
extracting applications (.sea)
REMARKS.....Compact Pro is NOT in this archive. Should be in
software archive.
SUB_PROJECT_OF.....RAMSES
CROSS_REFERENCES.....MacMETH, RAMSES, software archive
ARCHIVIST.....Frank Thommen
CREATION_DATE.....5.7.1996
MEDIA.....MO disks
LOCATION.....Treasor CHN C-stock / Schrank CHN D117
NAME.....SE 9 / SE10
```

Fig. 1. Beispiel eines **Metadaten** Eintrags für ein Archivpaket des Systemökologie Archivs. Die Metadaten wurden immer mit den gleichen Feldern gemäss einer Vorlage erstellt.

Datenmanagement wurde als **Führungsaufgabe** betrachtet und es gab ein "Drehbuch" zur Erstellung eines Archiveintrags. Darin waren Formate und Attribute festgelegt, es gab **Vorlagen** für Metadaten (Fig. 1), die zentral verwaltet wurden und Datenträger wurden nach bestimmten Zeitabständen umkopiert mit dem Ziel, die **Lesbarkeit und Wiederverwendbarkeit der Daten** zu gewährleisten.

3 Was sich bewährt hat

- **Risikominimierung** durch absichtliche **Redundanz**
 - Texte immer in mehreren Formaten abspeichern (DOC, TXT, PDF).
 - Daten als ASCII-Text-Files speichern.
- **Open/Free**-Lösungen sind **Closed/Proprietary**-Lösungen vorzuziehen
 - Freie Software und Formate sind auch nach Jahrzehnten leichter lesbar.
 - Wiederverwendbarkeit von Daten ist somit gewährleistet.
- **Forschungsdatenmanagement** ist **Führungsaufgabe**
 - Forschungsdatenmanagement als Chefsache zu betrachten, hat sich bewährt.
 - Dank regelmässiger **Qualitätskontrolle** konnten die Lesbarkeit und Wiederverwendbarkeit von Forschungsdaten während fast drei Jahrzehnten sichergestellt werden.
 - Die in Datenmanagement ausgebildeten Generationen von Doktoranden und Postdoktoranden agierten als **Multiplikatoren** und bildeten ihrerseits neue Generationen von Forschenden darin aus.
 - Die sauberen Forschungsdaten konnten schneller und einfacher den Forschungspartnern **Open Access** zur Verfügung gestellt werden und trugen somit zu einer verbesserten internationalen **Zusammenarbeit** und **Bekanntheitsgrad** der Gruppe bei.
- **Aufbewahrung** von funktionstüchtiger **historischer Hardware und Software**
 - Beim Umsteigen auf neue Systeme wurde immer ein Exemplar funktionierender Hardware mit der dazugehörigen Software physisch archiviert.
 - Dies erwies sich als nützlich beim Lesen alter **Datenträger** oder der Umwandlung historischer in moderne **Fileformate**.

4 Probleme und Denkanstösse

- **Lesbarkeit** der Daten gewährleisten
 - Wenn Hard- und Software, die zum Lesen nötig waren, nicht mitarchiviert wurde, hatte dies schwerwiegende Datenverluste zur Folge.
 - **Materialalterung** der Hardware zerstörte Datenträger (Fig. 2).
 - Restaurationsarbeiten waren erschwert, da die Hardware **Spezifikationen** nicht **Open Access** waren.
- **Kontext** der Daten
 - Es genügt nicht, die Rohdaten zu archivieren und verfügbar zu stellen. Der Datenkontext muss ebenfalls **Open Access** sein.
- **Copyright**
 - Es wurde häufig versäumt, die **Einwilligung der Autoren** zur Open-Access-Publikation ihrer Schriften und Daten zu ihrer Lebzeit einzuholen.
 - Viele einzigartige Publikationen blieben **ungewollt** (auch vom verstorbenen Urheber) unter Verschluss.



Fig. 2. Beispiel eines durch **Materialalterung** beim Leseversuch zerstörten Datenträgers.

5 Schlussfolgerungen

- Wir empfehlen allen Forschenden, die um Open Access bemüht sind, als einen ersten Schritt ein bewusstes **Datenmanagement** zu betreiben und dies, falls sie in leitender Funktion tätig sind, aktiv bei ihren Mitarbeitenden zu fördern.
- Der **Datenkontext** soll genauso wie die Rohdaten Open Access verfügbar gemacht werden.
- Man muss nicht alle Daten und Publikationen auf einmal völlig Open Access stellen, sondern kann **Schrittweise** mit einzelnen Projekten unter **Creative-Commons-Lizenz** anfangen. Lieber klein anfangen, als gar keinen Open Access betreiben. Offene Lösungen zahlen sich auf die Dauer aus.