



Digitales Datenmanagement seit 1988

Überführung eines digitalen Forschungsarchivs aus dem Bereich Systemökologie in das ETH Data Archive

Dr. Ana Sesartić

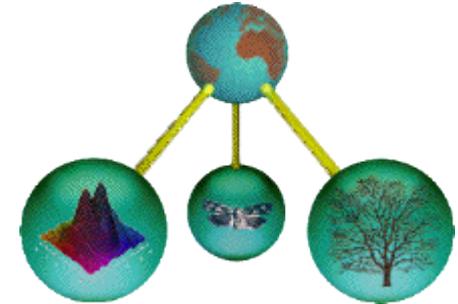
Digitaler Datenerhalt, ETH-Bibliothek, ETH Zürich

Zeitreise von



zu

- 1980er Systemökologie



- 1988-2015 Archiv



- Heute Überführung ETH Data-Archive

Rosetta

Systemökologie

*“We are an **interdisciplinary** group studying **dynamics** of **terrestrial ecosystems** as impacted by **environmental changes** such as climatic change.”*

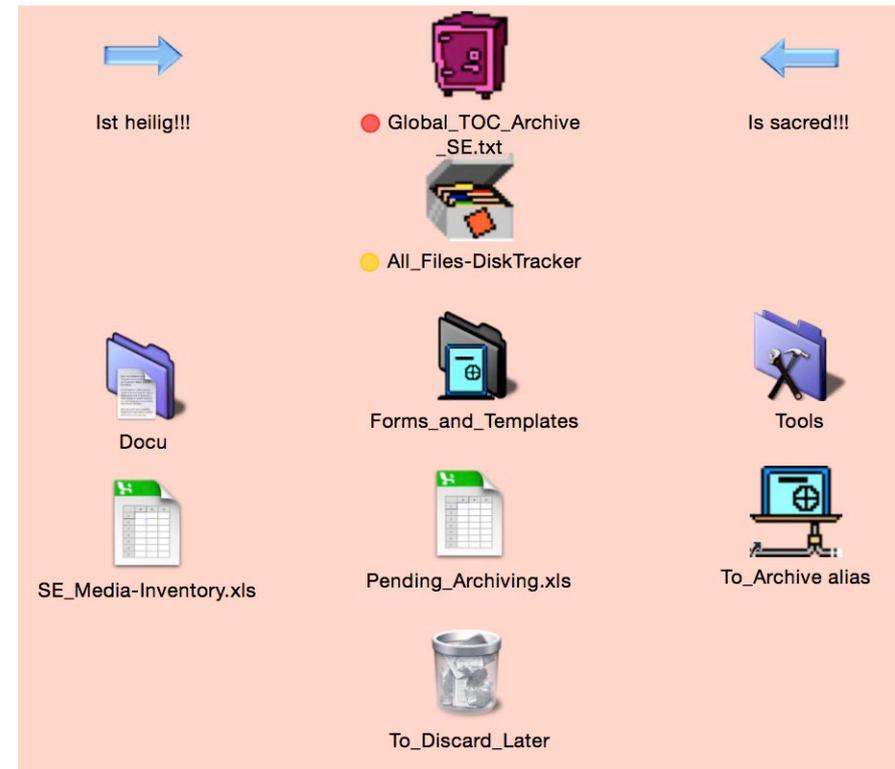
- Einfluss des Klimawandels auf Ökosysteme
- Systemtheorie und Modellierung
- Fachgruppe «Terrestrial Systems Ecology»
gegründet 1988 durch Prof. Andreas Fischlin
<http://www.sysecol.ethz.ch/>



© Andreas Fischlin

Archivkonzept

- «Drehbuch» zur Erstellung eines Archiveintrags
- Festgelegte Formate und Attribute
- Vorlagen für Metadaten Beschreibung
- Sammelverzeichnis für Metadaten
- Ziel: Lesbarkeit und Wiederverwendbarkeit von Daten



© SE Archive

Beispiel: «Drehbuch» für neuen Eintrag

1. Absprache des geplanten Eintrags mit Gruppenleiter
2. Vorbereitung der Archivelemente durch Ordnen und Wegwerfen obsolet gewordener Elemente
 - Text und Daten in mehreren Formaten speichern (XLSX als CSV und SLK, DOC als RTF und PDF)
 - Bei Bedarf zum Lesen der Files nötige Software mitarchivieren
3. Besorgen der erforderlichen Archivmedien
4. Erstellen der Eintragsbeschreibung (ASCII Text)
5. Ablegen der Eintragsbeschreibung
6. Ablegen des Archiveintrags (eigentliche Archivierung, in doppelter Ausführung)
7. Erstellen der Medieneintragsbeschreibung
8. Nachführen der globalen Dokumentation
9. Abschluss (Archivgut lagern, je eine Kopie in Tresor und in Archivschrank)

Metadaten

- TITLE
- VERSION
- LIST_OF_PARTS
- AUTHORS
- DESCRIPTION
- HANDLING
- REMARKS
- SUB_PROJECT_OF
- CROSS_REFERENCES
- ARCHIVIST
- CREATION_DATE
- MEDIA
- LOCATION
- NAME

```

TITLE.....RAMSES 2.2b6 (baikal)

VERSION.....2.2b6

LIST_OF_PARTS.....:00_Index (/pub/mac/RAMSES/)
                   :00_README.RAMSES
                   :MManual_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.ito.umnw.ethz.ch/pub/mac/RAMSES/"
                   "(=baikal.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.....Tresor CHN C-stock / Schrank CHN D117

NAME.....SE 9 / SE10

```

Archiv – Inhalte und Formate

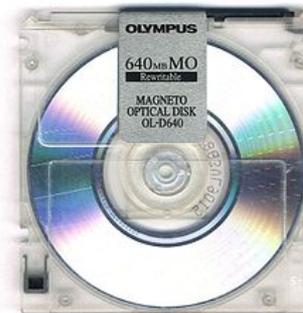
- Publikationen
 - DOC, RTF, TXT, TEX, PDF
- Grafiken
 - PNG, JPEG, TIFF, PICT
- Software
 - MODULA, R, SHELL, AppleScript
- Daten
 - CSV, DAT, TXT, XLS, SLK
- Pakete
 - StuffIt, TAR, ZIP
- Absichtliche Redundanz
 - Eines der Formate in Zukunft hoffentlich lesbar
 - Wiederverwendbarkeit gewährleistet
 - Risikominimierung
- Generelle Richtlinien
 - OpenSource > ClosedSource
 - Free > Proprietary
 - Standards > Obskures

Speichermedien

- 1980er Magnetbänder
 - 1990er Magnetoptische Speicher
 - 1990er & 2000er CD
 - 2010er DVD
-
- Umkopieren der Daten auf die jeweils aktuelle Trägertechnologie gemäss vorgegebenem Zeitplan...
 - ...mit Hindernissen



© SE Archive



© Wikipedia

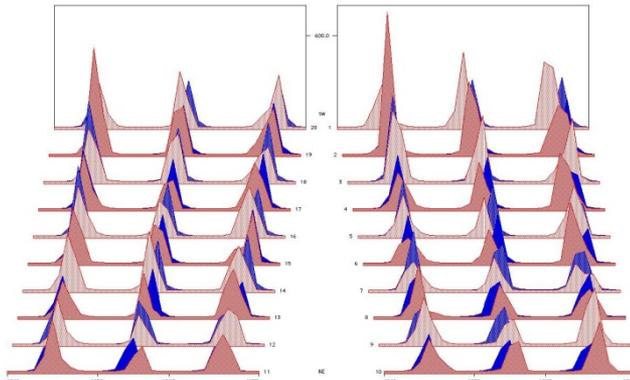


© Wikipedia

Herausforderungen bei der Überführung in ein digitales Langzeitarchivierungssystem

Situation

- Einzigartige Datenbank mit 63 Jahren ökologischer Beobachtungsdaten
 - 3 Professuren
 - Mehrere Mio. Forschungsgelder
 - Aktuelle Relevanz (Klimawandel)



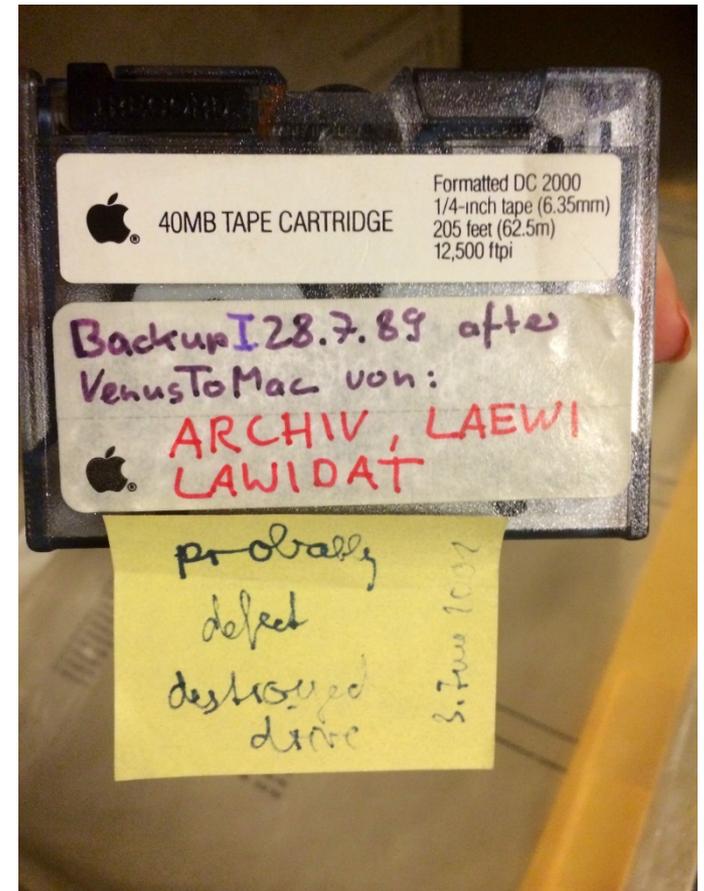
Fischlin (1979, 1982), Baltensweiler & Fischlin (1988)

Fragen

- Was macht man, wenn die zum Lesen nötige Hardware nicht mehr vorhanden ist?
- Wie lösen wir Kompatibilitäts- und Encoding Probleme?
- Wie sichern wir die Lesbarkeit der verbliebenen Daten?

Stolpersteine, Denkanstösse und Lerneffekte

- Auch «alte» digitale Daten sind archivierungswürdig und können sich als höchstrelevant erweisen
- Reine ASCII Files erweisen sich als am robustesten
- Nicht-ASCII Encodings (z.B. Umlaute) können ganze Datensätze unbrauchbar machen
- Formate müssen regelmässig auf die neuste Version gehoben werden
- Software muss mit-archiviert werden
- Funktionstüchtige Hardware müsste ebenfalls archiviert werden



© SE Archive

Baustellen für die Community

- Archivierung von Software
- Archivierung von Hardware
- Überwindung der UNIX/Mac/Windows Gräben
- Stärkere Zusammenarbeit zwischen Forschung, Informatik, Bibliotheken und Archiven

Die beste Langzeitarchivierung für Files ist dann gewährleistet, wenn mit ihnen gearbeitet wird.

Fragen?

www.library.ethz.ch/Digitaler-Datenerhalt

data-archive@library.ethz.ch

Dr. Ana Sesartić
ETH-Bibliothek
Digitaler Datenerhalt
Rämistrasse 101
8092 Zürich
+41 44 632 73 76
ana.sesartic@library.ethz.ch