

Forschungsdateninfrastruktur und Langzeitarchivierung

daherb

GPN21, 20230608

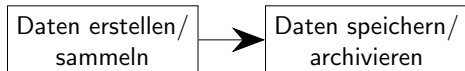
Über mich



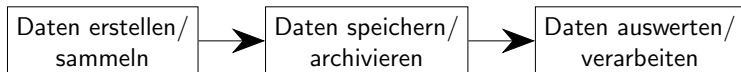
Forschungsdatenkreislauf

Daten erstellen/
sammeln

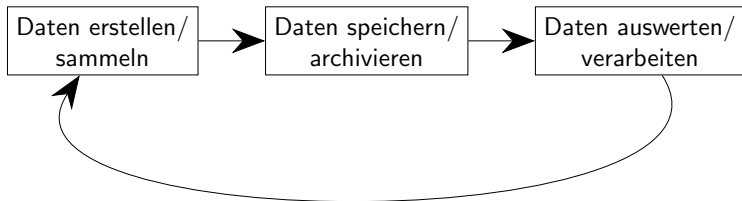
Forschungsdatenkreislauf



Forschungsdatenkreislauf



Forschungsdatenkreislauf



Forschungsdaten



<https://commons.wikimedia.org/wiki/File:Zettelarchiv2.JPG> N p holmes, CC BY-SA 3.0, via Wikimedia Commons

Forschungsdaten



[https://digitaltmuseum.org/021018072714/tele-och-dataanlaggning Järnvägsmuseet, unbekannter Fotograf, Public domain](https://digitaltmuseum.org/021018072714/tele-och-dataanlaggning_Järnvägsmuseet,_obekannter_Fotograf,_Public_domain)

Forschungsdaten

P 02 12 CLAUSIT		CLAUSIT		430 035 043	
STIPENDIA	VERBIS	SCHEDA PRIMA			
STIPENDIA	VERBIS	SCHEDA VERSUUM		LEMMATA	VERBORUM
STIPENDIA	VERBIS	SCHEDA LEMMATUM		LEMMATA	VERBORUM
STIPENDIA	VERBIS	SCHEDA VERBORUM		LEMMATA	VERBORUM
STIPENDIA	VERBIS	SCHEDA VERBORUM A CONTRARIO ORD.		LEMMATA	VERBORUM

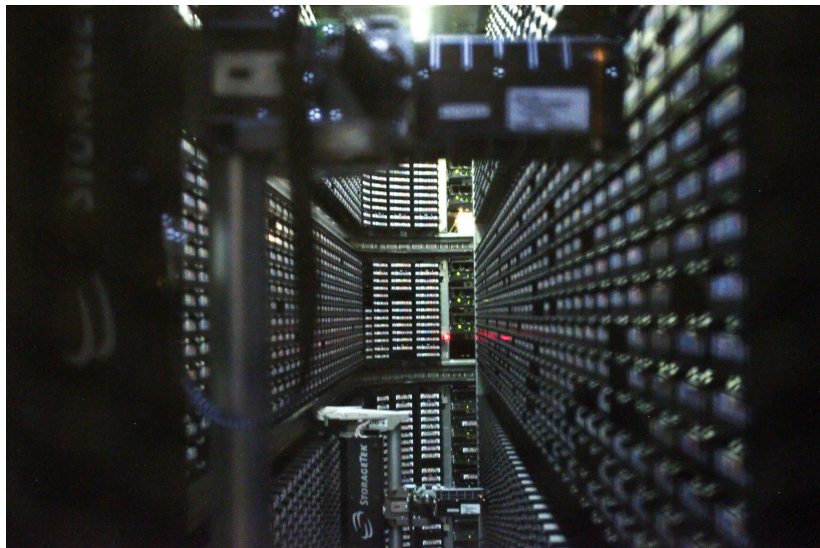
<https://umanisticadigitale.unibo.it/article/download/8575/9233/30311> Copyright (c) 2019 Geoffrey Rockwell, Marco Passarotti, CC-BY 4.0

Forschungsdaten



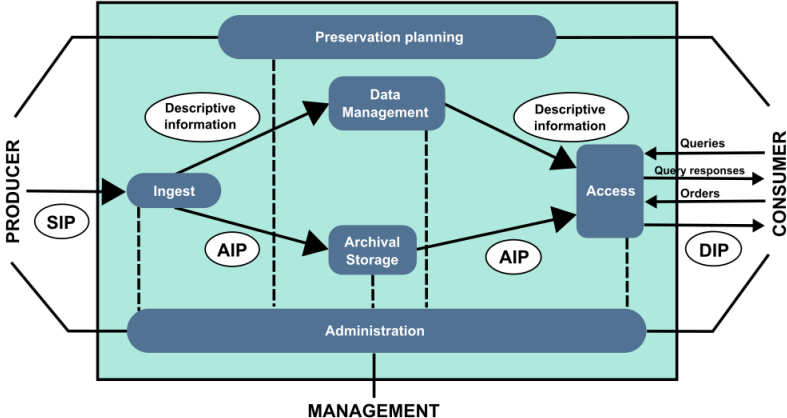
<https://www.pxfuel.com/en/free-photo-xpnug> Free for commercial use

Forschungsdaten

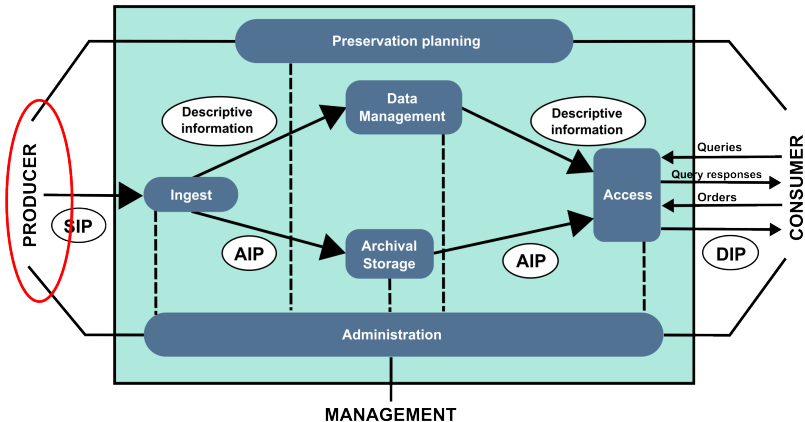


[https://commons.wikimedia.org/wiki/File:Interior_of_StorageTek_tape_library_at_NERSC_\(1\).jpg](https://commons.wikimedia.org/wiki/File:Interior_of_StorageTek_tape_library_at_NERSC_(1).jpg)
CC0, via Wikimedia Commons

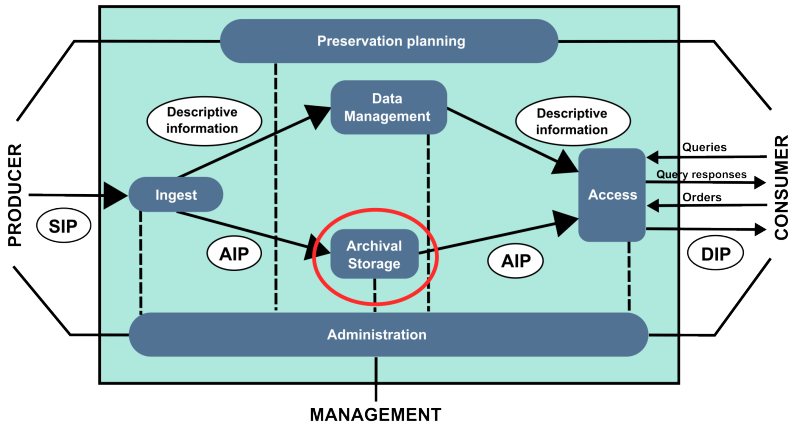
OAIS



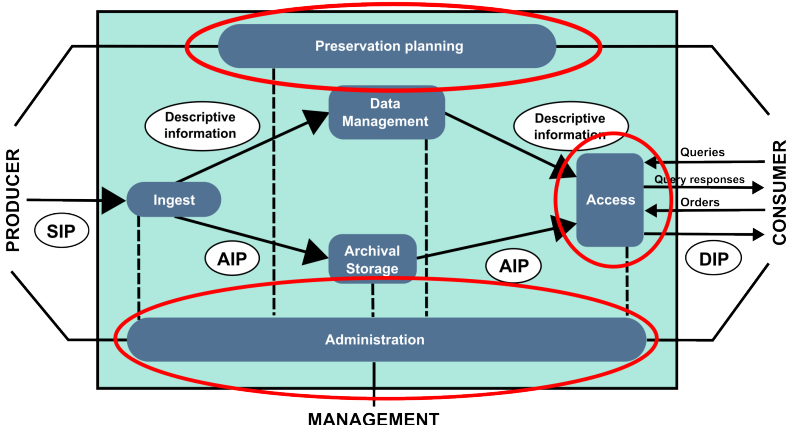
[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons



[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons

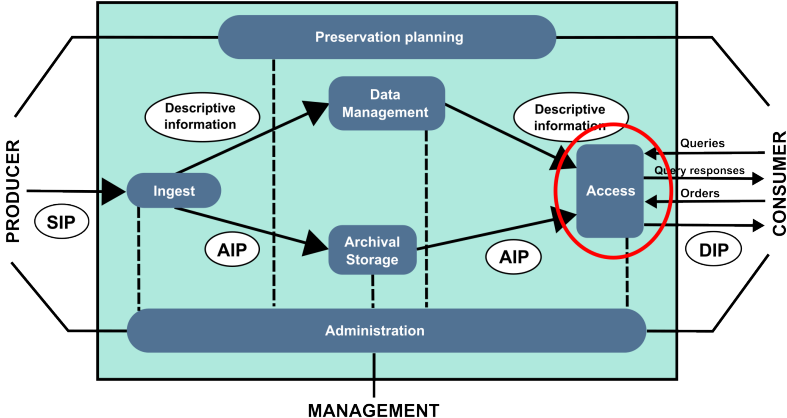


[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons



[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons

OAIS



[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons

Forschungsinfrastruktur

Subject ▾

Country ▲

Type to filter or search for more

Germany (26679)
Romania (1790)
Austria (1165)
United States (525)
Switzerland (349)
Poland (315)
Czech Republic (236)
Netherlands (116)
Estonia (111)
Polen (99)
more...

Organisation ▾

Data provider ▲

Type to filter or search for more

Bayerisches Archiv für Sprachsignale (16398)
Leibniz-Institut für Deutsche Sprache (14267)
MPI for Psycholinguistics (7451)
Other (5308)
Berlin-Brandenburg Academy of Sciences and Humanities (5059)
Meertens Instituut/Huic (759)
DARIAH-DE Repository (449)
LINDAT/CLARIAH-CZ (226)
Eberhard Karls Universität Tübingen (143)
Georg Eckert Institute for International Textbook Research (127)
more...

National project ▾

▲

It comprises a collection of user queries to a naturally spoken Web interface with the main focus on the soccer world series in 2006. The recordings include 156 field recordings using a hand-held U...

English German

📄 Landing page for this record

🔍 VCR

SmartKom Public

(Part of Bavarian Archive for Speech Signals (BAS))

📄 This corpus contains multi modal recordings of 86 actors who use the SmartKom system. SmartKom Public is comparable to a traditional public phone booth but equipped with additional intelligent communication devices. Naive users were asked to test a 'prototype' for a market study not knowing that the system was in fact ...

German English

📄 Landing page for this record

🔍 172 🔍 1 🗨️

🔍 VCR

Dissertation Data Dr. Veronika Neumeyer: Consonant Cluster Production in Cochlear Implant Patients

(Part of Bavarian Archive for Speech Signals (BAS))

📄 The Cl_2 corpora contain German speech recordings of 48 cochlear implant users (CI) and 48 speakers without hearing impairment (control group, KG). The data were analyzed in Veronika Neumeyer's dissertation "Akustische Analysen der Sprachproduktion von CI-Trägern" (2015). Cl_2_Cluster contains recordings used for the a...

English German

📄 Landing page for this record

🔍 36 🔍 1 🗨️

🔍 17 🗨️

🔍 VCR

The Zurich Tangram Corpus - BAS Edition

(Part of Bavarian Archive for Speech Signals (BAS))

📄 This corpus contains tasks, where one subject (the instructor) describes different Tangram figures to another subject (the receiver) so that the receiver can recreate the same order of figures that the instructor has in front of them. The subjects initially don't know each other and work together to solve these tasks ...

English German

📄 Landing page for this record

🔍 36 🔍 1 🗨️

🔍 17 🗨️

🔍 VCR

The Zurich Tangram Corpus - UZH Edition

(Part of Bavarian Archive for Speech Signals (BAS))

📄 This corpus contains tasks, where one subject (the instructor) describes different Tangram figures to another subject (the receiver) so that the receiver can recreate the same order of figures that the instructor has in front of them. The subjects initially don't know each other and work together to solve these tasks ...

English German

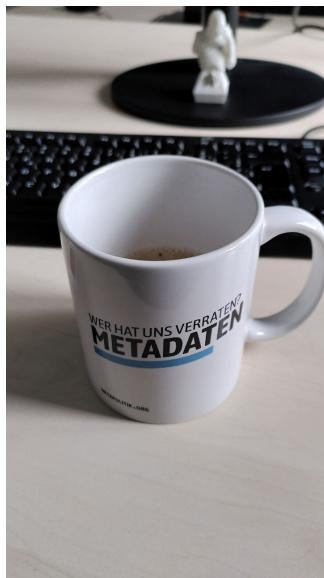
📄 Landing page for this record

🔍 36 🔍 1 🗨️

🔍 17 🗨️

🔍 VCR

Metadaten



Eigenes Bild

Metadaten

Metadaten

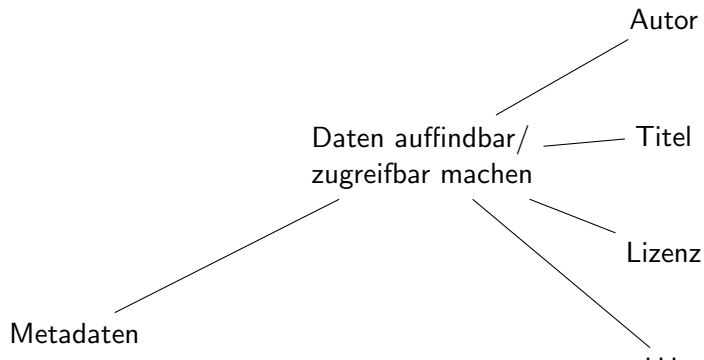
Metadaten

Daten auffindbar/
zugreifbar machen

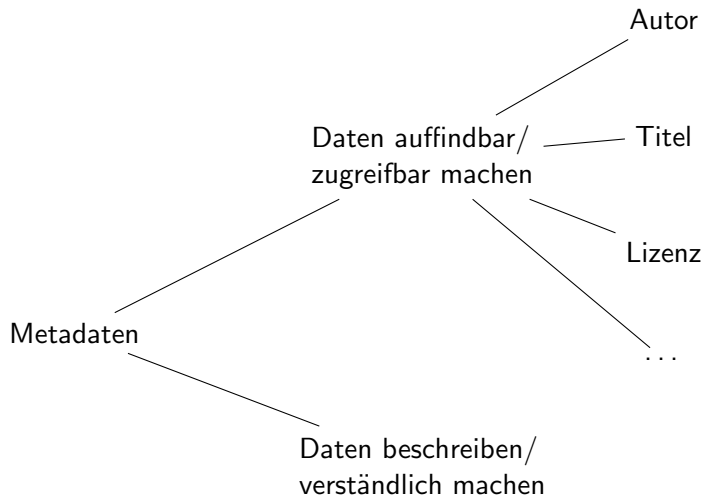
Metadaten



Metadaten



Metadaten



FAIR

Findable Metadaten und Daten sollten sowohl für Menschen als auch für Computer einfach auffindbar sein

Accessible Wenn User die benötigten Daten gefunden haben muss bekannt sein wie auf die Daten zugegriffen werden kann

Interoperable Daten müssen üblicherweise mit anderen Daten verknüpft werden

Reusable Das ultimative Ziel von FAIR ist die Wiederverwendung von Daten zu optimieren

Quelle: <https://www.go-fair.org/fair-principles/> Abgerufen 5.6.2023

Daten verstehen

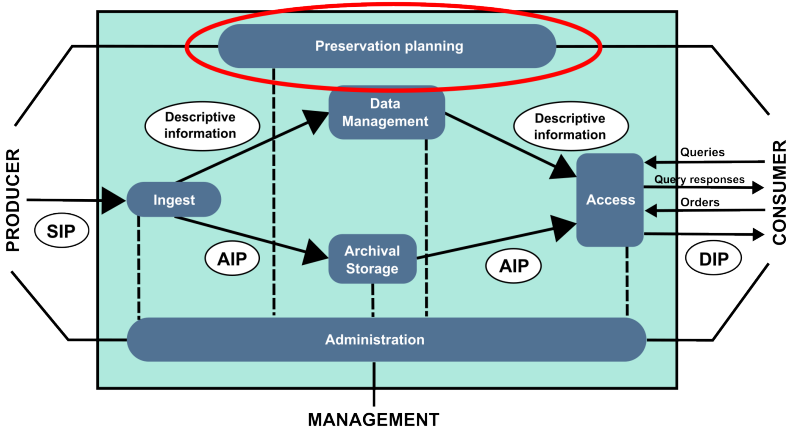
4530030 100

```
the -0.0811 0.5145 0.0368 0.0544 -0.0662 0.2121 0.1636 -0.1011 0.0607 -0.1858 -0.1965 -0.0595 0.0703 \
-0.1013 0.3130 -0.1717 0.0847 -0.1222 0.1024 0.0753 -0.1384 0.0435 -0.0371 0.1932 -0.1226 -0.2227 \
-0.1530 -0.2890 0.2371 0.2699 0.2693 -0.1666 0.0240 0.1053 -0.1475 -0.3232 0.0236 -0.2056 0.2847 \
-0.2817 0.1197 0.0314 -0.1215 0.0782 -0.2850 -0.1316 -0.0844 0.1483 -0.2192 -0.0462 -0.2151 0.0582 \
0.0372 0.0127 -0.3074 -0.1582 -0.1393 0.0361 -0.2519 -0.0305 -0.1532 -0.0286 -0.0955 0.3037 0.5632 \
-0.1120 -0.0319 -0.2223 -0.2612 -0.2254 -0.1593 0.1807 0.1205 0.3695 -0.2652 -0.0490 -0.2556 0.0130 \
-0.0898 0.0322 0.0021 -0.2692 0.3129 0.0179 0.3913 0.5415 -0.0049 0.0884 0.1605 0.0878 0.0004 0.1465 \
0.1872 0.0521 -0.1492 -0.0882 0.1696 0.1894 -0.0866 0.1184 \
in 0.1245 0.4200 0.2936 0.0924 -0.0669 0.0252 0.1407 -0.0729 0.0680 -0.2951 -0.2720 0.0785 0.0780 \
0.0248 0.0427 -0.1497 0.1013 -0.0257 0.0364 0.2647 0.0330 0.1047 0.0382 0.0138 -0.0162 -0.0733 0.0960 \
-0.2090 0.0561 0.1030 0.2898 -0.1914 -0.0927 0.1237 -0.0023 -0.4792 0.0523 -0.0819 0.3551 -0.2274 \
0.3301 0.0547 -0.1707 0.2304 -0.2599 -0.1389 -0.0106 0.1921 -0.3615 0.0077 -0.2439 0.1056 -0.0010 \
-0.2522 -0.2321 -0.1604 -0.2652 -0.0134 -0.3000 0.1215 0.0737 0.0215 -0.1647 0.2799 0.5886 -0.0189 \
-0.1250 -0.2438 -0.1621 -0.3960 -0.1078 0.2162 -0.1173 0.6267 -0.0788 0.0086 -0.1317 0.1440 -0.2035 \
-0.0742 -0.0536 -0.2773 0.1026 0.0356 0.3384 0.4606 -0.1893 0.0142 0.2269 0.1136 0.0962 0.1648 0.1065 \
0.0051 -0.1777 -0.0971 0.3434 0.2423 -0.1432 -0.1808 \
of 0.0018 0.3950 0.0662 0.1279 -0.2606 0.3275 -0.0161 -0.1740 0.0530 -0.1168 -0.1781 0.0897 0.0423 \
-0.1886 0.2358 -0.1118 0.1510 -0.2601 0.0367 -0.0689 -0.1016 0.0333 0.0290 0.1104 -0.3067 -0.2721 \
-0.1113 -0.2921 0.4445 0.1993 0.2558 -0.1089 0.0121 0.0870 -0.2314 -0.2405 -0.0331 -0.5395 0.1571 \
-0.3647 0.1048 0.0985 -0.1880 -0.0555 -0.0523 -0.0162 -0.0134 0.0990 -0.1943 0.1087 -0.3270 0.0789 \
-0.0764 0.1340 -0.2139 -0.1241 -0.3107 -0.2037 -0.3198 0.0157 0.0148 -0.0511 -0.1739 0.4170 0.6791 \
-0.1775 -0.1529 -0.1850 -0.2302 -0.3126 -0.0604 0.2185 0.0143 0.5491 -0.1606 0.0136 -0.2309 0.1908 \
-0.2026 0.0611 -0.1822 -0.3973 0.2603 0.1247 0.4455 0.4626 -0.0074 -0.0384 0.1773 0.1086 0.1534 0.2286 \
0.1497 -0.0698 -0.0397 -0.0486 0.0400 0.1595 -0.0070 0.2451 \
a -0.1433 0.4563 0.0432 0.2349 -0.2987 0.0416 0.0030 -0.0698 -0.0150 -0.1932 -0.0465 -0.0910 0.2336 \
-0.1313 0.2681 -0.0743 0.1192 -0.0902 -0.1458 0.2554 0.0500 -0.0215 -0.2235 0.3174 -0.0591 -0.2207 \
-0.2203 -0.2378 0.0418 0.3202 0.2689 -0.2710 -0.0955 0.2005 -0.1985 -0.3306 -0.2077 -0.2333 0.1841 \
-0.2972 0.0294 -0.0136 -0.0701 0.1860 -0.1163 0.1043 0.0794 0.1794 -0.3520 0.0179 -0.3910 0.0755 0.0074 \
0.0377 -0.1803 -0.0884 -0.2182 0.0964 -0.1289 0.2162 -0.3919 -0.0599 -0.2586 0.0642 0.4901 0.0188 \
-0.1292 -0.3284 -0.0449 -0.2160 0.0986 0.3132 0.1201 0.3162 -0.2393 0.1444 -0.2775 0.2539 -0.0494 \
0.1743 -0.0795 -0.1996 0.2118 -0.2815 0.3230 0.4166 0.1315 0.2142 -0.1657 -0.0734 -0.0965 0.0446 0.1554 \
-0.0070 -0.0130 -0.3514 0.1540 0.2946 -0.3259 0.0810
```


Daten verstehen

Besser:

- ▶ Aussagekräftige Header
- ▶ Kontrollierte Vokabulare
- ▶ Menschen- und Maschinenlesbar
- ▶ Etablierte Standards
- ▶ Validierbare Formate



[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons

Konservierungsrichtlinien

Beispiel: Flash

Digitale Animation als Flash Video (flv)
(Unsupported seit 12. Januar 2021)

Konservierungsrichtlinien

Beispiel: Flash

Digitale Animation als Flash Video (flv)
(Unsupported seit 12. Januar 2021)

Konvertierung

MPEG-4

Konservierungsrichtlinien

Beispiel: Flash

Digitale Animation als Flash Video (flv)
(Unsupported seit 12. Januar 2021)

MPEG-4

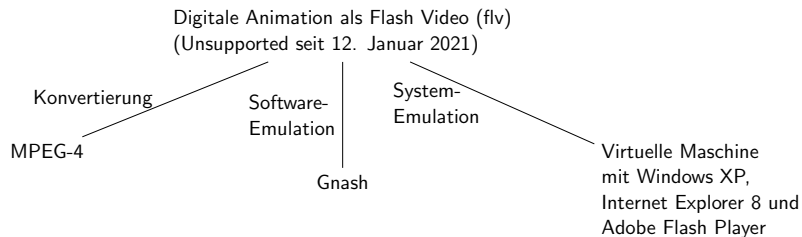
Konvertierung

Software-
Emulation

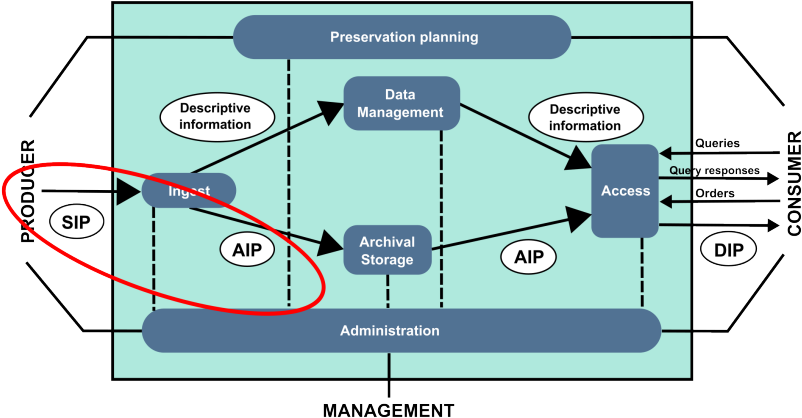
Gnash

Konservierungsrichtlinien

Beispiel: Flash



OAIS



[https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_\(en\).svg](https://en.wikipedia.org/wiki/Open_Archival_Information_System#/media/File:OAIS_Functional_Model_(en).svg) Mathieu-Alex Haché (original work), Mess (SVG conversion and English translation), CC-BY-SA 4.0, via Wikimedia Commons

Informationspakete

Übermittlungspaket

sip-bag

|--data

| |-- Metadata

| | |-- Catalog Metadata (CMDI) (1..)

| | \-- ACL (1) (?)

| |-- Resources

| | |-- Schemas (1..) (?)

| | \-- License (1..)

| \-- Content

| | \-- File (1..)

|-- manifest-{checksum}.txt

\-- bagit.txt

Informationspakete

Archivpaket

aip-bag

|--data

| |-- Metadata

| | |-- Catalog Metadata (CMDI) (1..)

| | |-- Archival Metadata (z.B. Premis in Mets) (1)

| | \-- ACL (1)

| |-- Resources

| | |-- Schemas (1..)

| | \-- License (1..)

| |-- Bitstream

| | \-- File (1..)

| \-- Converted

| \-- File (0..)

|-- manifest-{checksum}.txt

\-- bagit.txt

HEY, LOOK, WE HAVE A BUNCH
OF DATA! I'M GONNA ANALYZE IT.

NO, YOU FOOL! *THAT WILL
ONLY CREATE MORE DATA!*



<https://xkcd.com/2582/> Randall Munroe, CC-BY-NC 2.5

It's important to make sure your analysis destroys as much information as it produces.