
DEAL Dark Archive an der TIB

Ein Einblick in die Langzeitarchivierung

DEAL Praxis Webinar
29. Oktober 2024, 10 Uhr, Online
Micky Lindlar

Agenda

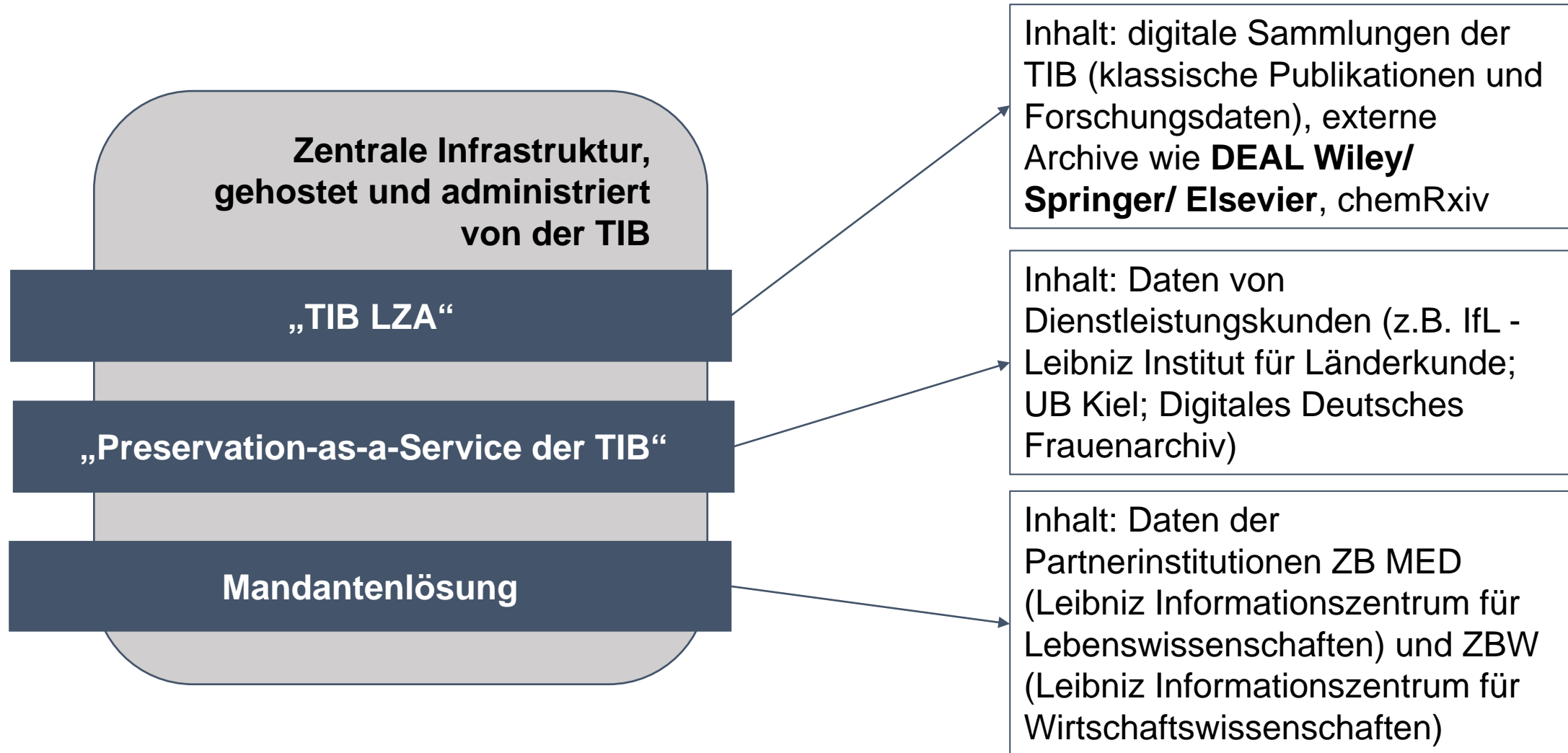
- Zum Hintergrund: Digitale Langzeitarchivierung an der TIB
- Was ist ein Dark Archive?
- Wann darf Zugriff gegeben werden?
- Was wird archiviert? (oder: Was ist ein Artikel?)
- Ist auch alles da?
- IST-Stand Archive

Digitale Langzeitarchivierung an der TIB

- Aktiv in der Digitalen Langzeitarchivierung seit 2009, Produktivbetrieb seit 2015
- Regelmäßige Zertifizierungen als **vertrauenswürdigen digitales Langzeitarchiv** (DataSeal of Approval 2015, nestor Siegel 2017, CoreTrustSeal 2020, nestor Siegel 2022, aktuell laufend: CoreTrustSeal Re-Zertifizierung)
- **Personelle Ressourcen:**
6,4 Vollzeitäquivalente (8 Personen) im Team LZA + 1,6 Vollzeitäquivalente (6 Personen) für Aufgabe LZA ausserhalb des Teams LZA (anteilige Stellen aus Erwerbungssteams, anteilige Stellen aus IT)
- **Infrastrukturelle Ressourcen:**
3 Systemumgebungen (Entwicklungssystem, Testsystem, Produktivsystem); Produktivsystem mit 4 Application Servern + db-Server; Softwarekern: Rosetta von Ex Libris; aktuell 2 getrennte Kopien, dritte für 2025 geplant
- **Netzwerke:**
Aktiv in nationalen und internationalen LZA-Netzwerken
- Weitere Infos unter:
<https://www.tib.eu/de/publizieren-archivieren/digitale-langzeitarchivierung>



TIB LZA– 3 Rollen in einer Infrastruktur



Was bedeutet „Dark Archive“?

Light / Dim / Dark Archive beschreibt die **direkte Anbindung** der Zugriffsplattform für Nutzende an die Archivinfrastruktur / den Archivspeicher

Light = Nutzungskopie wird immer direkt aus dem Archiv/ Repositorym geliefert (Beispiel: klassische Publikationsplattformen)

Dim = manche, aber nicht alle Nutzungskopien werden direkt aus dem Archiv/ Repositorym geliefert

Dark = Nutzungskopien werden nie direkt aus dem Archiv/ Repositorym geliefert sondern nur im „Trigger Event“ (z.B. Verlagsplattform für Zugriff)

→ Trennung von Archiv- und Nutzungsinfrastruktur erhöht Datensicherheit

→ Dark Archive heißt **nicht** „Black Box auf die niemand jemals zugreifen kann“



Trigger Events in DEAL Verträgen

Springer Nature:

*„If Publisher does not comply with its obligations as under Section 1.2 of the Publish and Read Agreement Terms and Conditions for a period of **30 days or more** after receiving written notice by DEAL Operating Entity of Publisher’s failure (including a detailed description of the circumstances of such failure) [...]“*

Quelle: Addendum Dark Archive, Section 1.2 in *Projekt DEAL - Springer Nature Publish and Read Agreement, 2024-2028*. doi:[10.17617/2.3551270](https://doi.org/10.17617/2.3551270).

Trigger Event:
Was muss passieren, damit Zugriff gegeben werden darf?

Abgedeckte Fälle:
- Verlagsplattform nicht mehr verfügbar (> 30 Tage)
- Konkurs (> 60 Tage)

Wiley:

*“if Wiley makes a general assignment for the benefit of creditors, or files a voluntary petition in bankruptcy [...] within **60 days** after the filing [...]; or downtime of Wiley Online Library and/or a total loss of access by Member Institution to the Licensed Electronic Products, either of which are continuing **for at least 30 days**.”*

Quelle: Section 3.6.3 in *Projekt DEAL - Wiley Publish and Access Agreement, 2024-2028*. doi:[10.17617/2.3551268](https://doi.org/10.17617/2.3551268).

Elsevier:

*“(a) If Elsevier makes a general assignment for the benefit of creditors, or files a voluntary petition in bankruptcy [...] within **60 days** after the filing [...]; or (b) a total loss of access to the Products **for at least 30 days** caused by a failure of Elsevier’s systems.”*

Quelle: Section 3.6.3 in *Projekt DEAL – Elsevier Publish and Read Agreement*. doi:[10.17617/2.3523659](https://doi.org/10.17617/2.3523659).

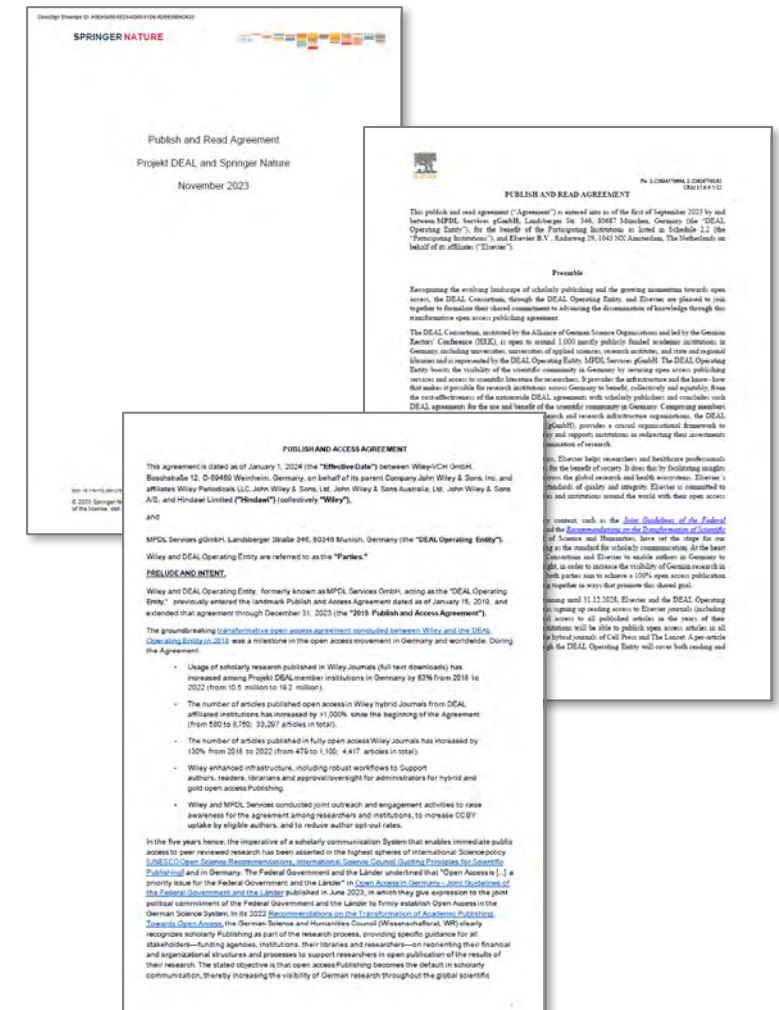
Anforderungen an Archivierung & Bereitstellung in Verträgen

Die DEAL Verträge regeln neben Trigger Events u.a.:

- Sicherheitsanspruch an Dark Archive Entity
- Access nur durch Konsortialteilnehmer
- Nachweispflicht über Zugriffe im Triggerevent (z.B. Springer Nature: Nutzungsdaten für Volltext-Downloads nach Titel und/oder Institution min. jährlich)

Die DEAL Dark Archive Agreements:

- Sind abgeschlossen zwischen „Dark Archiving Entity“ (TIB) und dem Verlag
- Regeln Lieferweg (z.B. SFTP) und Lieferstruktur der Artikel
- Wiederholen Archivrelevante Bestandteile des Hauptvertrags (u.a. das Recht auf Migration „shall include the right to any technical modification [...] required to establish and/or maintain the Dark Archive“)



Was wird archiviert? (oder: Was ist ein Artikel?)



Cell Biol Toxicol (2023) 39:641–655
<https://doi.org/10.1007/s10565-021-09618-9>

ORIGINAL ARTICLE

Diesel exhaust exposure alters the expression of networks implicated in neurodegeneration in zebrafish brains

M. Saeid Jami · Hiromi Murata · Lisa M. Barnhill · Sharon Li · Jeff M. Bronstein 

Received: 5 December 2020 / Accepted: 12 May 2021 / Published online: 31 May 2021
© The Author(s) 2021

Abstract Neurodegenerative diseases are a major cause of disability in the world, but their etiologies largely remain elusive. Genetic factors can only account for a minority of risk for most of these disorders, suggesting environmental factors play a significant role in the development of these diseases. Prolonged exposure to air pollution has recently been identified to increase the risk of Alzheimer's and Parkinson's diseases, but the molecular mechanisms by which it acts are not well understood. Zebrafish embryos exposed to diesel exhaust particle extract (DEPe) lead to dysfunctional autophagy and neuronal loss. Here, we exposed zebrafish embryos to DEPe and performed high throughput proteomic and transcriptomic expression analyses from their brains to identify pathogenic pathways induced by air pollution. DEPe treatment altered several biological processes and signaling pathways relevant to neurodegenerative processes, including xenobiotic metabolism, phagosome maturation, and amyloid processing. The biggest induction of gene expression in brains was in Cyp1A (over 30-fold). The relevance of this expression change was confirmed by blocking induction using CRISPR/Cas9, which resulted in a dramatic increase in sensitivity to DEPe toxicity, confirming that Cyp1A induction was a compensatory protective mechanism. These studies identified disrupted molecular pathways that may contribute to the pathogenesis of neurodegenerative disorders. Ultimately, determining the molecular basis of how air pollution increases the risk of neurodegeneration will help in the development of disease-modifying therapies.

Highlights

- Air pollution increases the risk of developing neurodegenerative diseases.
- Zebrafish exposure to diesel exhaust particle extract (DEPe) results in dysfunction in autophagy and neuronal loss.
- Transcriptomic and proteomic alterations in zebrafish brains in response to exposure to DEPe reveals pathogenic pathways.
- The importance of increased Cyp1a expression is highlighted by an increased toxicity to DEPe after it is knocked down.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10565-021-09618-9>.

M. S. Jami · L. M. Barnhill · S. Li · J. M. Bronstein (✉)
Department of Neurology, David Geffen School of Medicine At UCLA, 710 Westwood Plaza, Los Angeles, CA 90095, USA
e-mail: jbronte@mednet.ucla.edu

H. Murata · L. M. Barnhill · J. M. Bronstein
Molecular Toxicology IDP, David Geffen School of Medicine At UCLA, Los Angeles, CA, USA

 Springer

Home > Cell Biology and Toxicology > Article

Diesel exhaust exposure alters the expression of networks implicated in neurodegeneration in zebrafish brains


Original Article | Open access | Published: 31 May 2021
Volume 39, pages 641–655, (2023) [Cite this article](#)

[Download PDF](#)  You have full access to this [open access](#) article

[Cell Biology and Toxicology](#)

[Aims and scope](#) →

[Submit manuscript](#) →


M. Saeid Jami, Hiromi Murata, Lisa M. Barnhill, Sharon Li & Jeff M. Bronstein 

 4058 Accesses  10 Citations  20 Altmetric  2 Mentions [Explore all metrics](#) →

Abstract

Neurodegenerative diseases are a major cause of disability in the world, but their etiologies largely remain elusive. Genetic factors can only account for a minority of risk for most of these disorders, suggesting environmental factors play a significant role in the development of these diseases. Prolonged exposure to air pollution has recently been identified to increase the risk of Alzheimer's and Parkinson's diseases, but the molecular mechanisms by which it acts are not well understood. Zebrafish embryos exposed to diesel exhaust particle extract (DEPe) lead to dysfunctional autophagy and neuronal loss. Here, we exposed zebrafish embryos to DEPe and performed high throughput proteomic and transcriptomic expression analyses from their brains to identify pathogenic pathways induced by air pollution. DEPe treatment altered several biological processes and signaling pathways relevant to neurodegenerative processes, including xenobiotic metabolism, phagosome maturation, and amyloid processing. The biggest induction of gene expression in brains was in Cyp1A (over 30-fold). The relevance of this expression change was confirmed by blocking induction using CRISPR/Cas9, which resulted in a dramatic increase

[Use our pre-submission checklist](#) →

Avoid common mistakes on your manuscript. 

Sections [Figures](#) [References](#)

[Abstract](#)

[Introduction](#)

[Material and methods](#)

[Results and discussion](#)

[Data Availability](#)

[References](#)

[Acknowledgements](#)

[Funding](#)

[Author information](#)

[Ethics declarations](#)

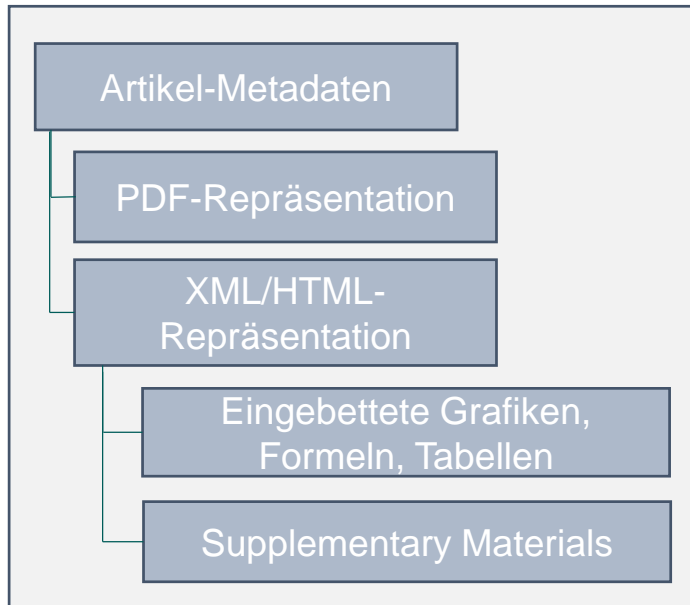
[Additional information](#)

[Supplementary Information](#)

Springer Testdatenlieferung
(PDF und Metadaten)

Darstellung des gleichen Artikels auf Springer Link
(siehe <https://doi.org/10.1007/s10565-021-09618-9>)

Vom Artikel zum Artikelobjekt



„Artikelobjekt“
mit Verlagen vereinbarter
Ablieferungsumfang pro Artikel

Home > Cell Biology and Toxicology > Article

Diesel exhaust exposure alters the expression of networks implicated in neurodegeneration in zebrafish brains

Original Article | Open access | Published: 31 May 2021
Volume 39, pages 641–655, (2023) | Cite this article

[Download PDF](#) **PDF Repräsentation**

M. Saeid Jami, Hiromi Murata, Lisa M. Barnhill, Sharon Li & Jeff M. Bronstein

4058 Accesses | 10 Citations | 20 Altmetric | 2 Mentions | [Explore all metrics](#)

The importance of increased Cyp1a expression is highlighted by an increased toxicity to DEPe after it is knocked down.

Supplementary Information

Below is the link to the electronic supplementary material.

- [Supplementary file1 \(XLSX 10 KB\)](#)
- [Supplementary file2 \(XLSX 1329 KB\)](#)
- [Supplementary file3 \(XLSX 152 KB\)](#)
- [Supplementary file4 \(XLSX 4167 KB\)](#)
- [Supplementary file5 \(XLSX 2356 KB\)](#)

Cell Biology and Toxicology

Aims and scope →
Submit manuscript →

Use our pre-submission checklist →
Avoid common mistakes on your manuscript.

Sections | **Figures**

- Abstract
- Introduction
- Material and methods
- Results and discussion
- Data Availability
- References
- Acknowledgements
- Funding
- Author information
- Ethics declarations
- Additional Information
- Supplementary Information**

eingebettete
Grafiken, Formeln,
Tabellen

Supplementary
Information

Artikelobjekt: Warum PDF und XML/HTML?



PDF Repräsentation – „Digitales Papier“

Vorteile:

- Festes Seitenlayout / volle Kontrolle über Darstellung
- Gut geeignet für Druck
- IdR keine Abhängigkeit zu Internet

Nachteile:

- Nicht für alle Viewer gleich gut geeignet (Mobile Devices)
- Keine Supplements
- Keine dynamischen Objekte (z.B. AV)
- Für Nutzungsszenarien wie Textmining / Extraktion von Formeln schlecht geeignet

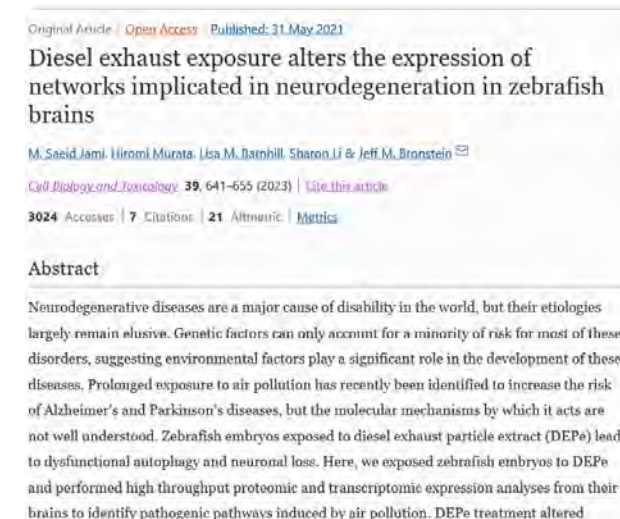
XML/HTML-Repräsentation –“Online Version“

Vorteile:

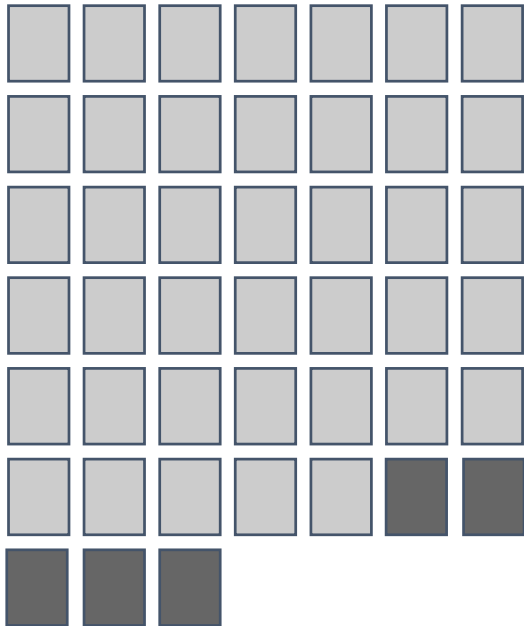
- Dynamisch anpassbar auf verschiedene Nutzungs/Viewerbedürfnisse
- Dynamische Inhalte wie Zitationsanalysen und Zugriffsstatistiken anzeigbar
- Supplementary Materials als Sidecar File aufrufbar

Nachteile:

- IdR nur Online verfügbar
- Schlecht ausdrückbar und u.U. abhängig von zusätzlichen Layoutinformationen



Datenmenge als Herausforderung: Beispiel Wiley Back Issues



Erwartete Daten laut Inventory-File:

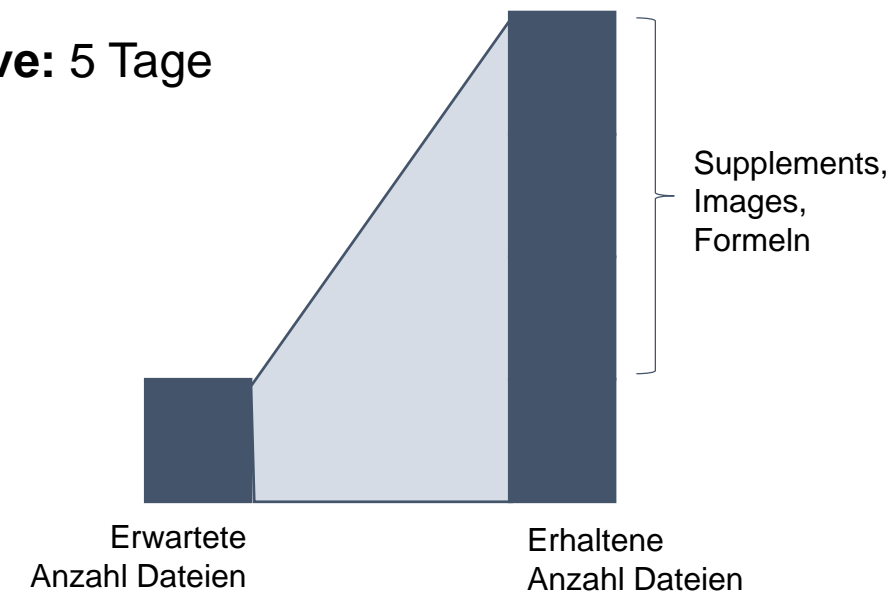
5.2 Millionen Artikel mit
12.8 Millionen Dateien
(geliefert in zip-Archiven)

Dauer Übertragung via SFTP: 40 Tage

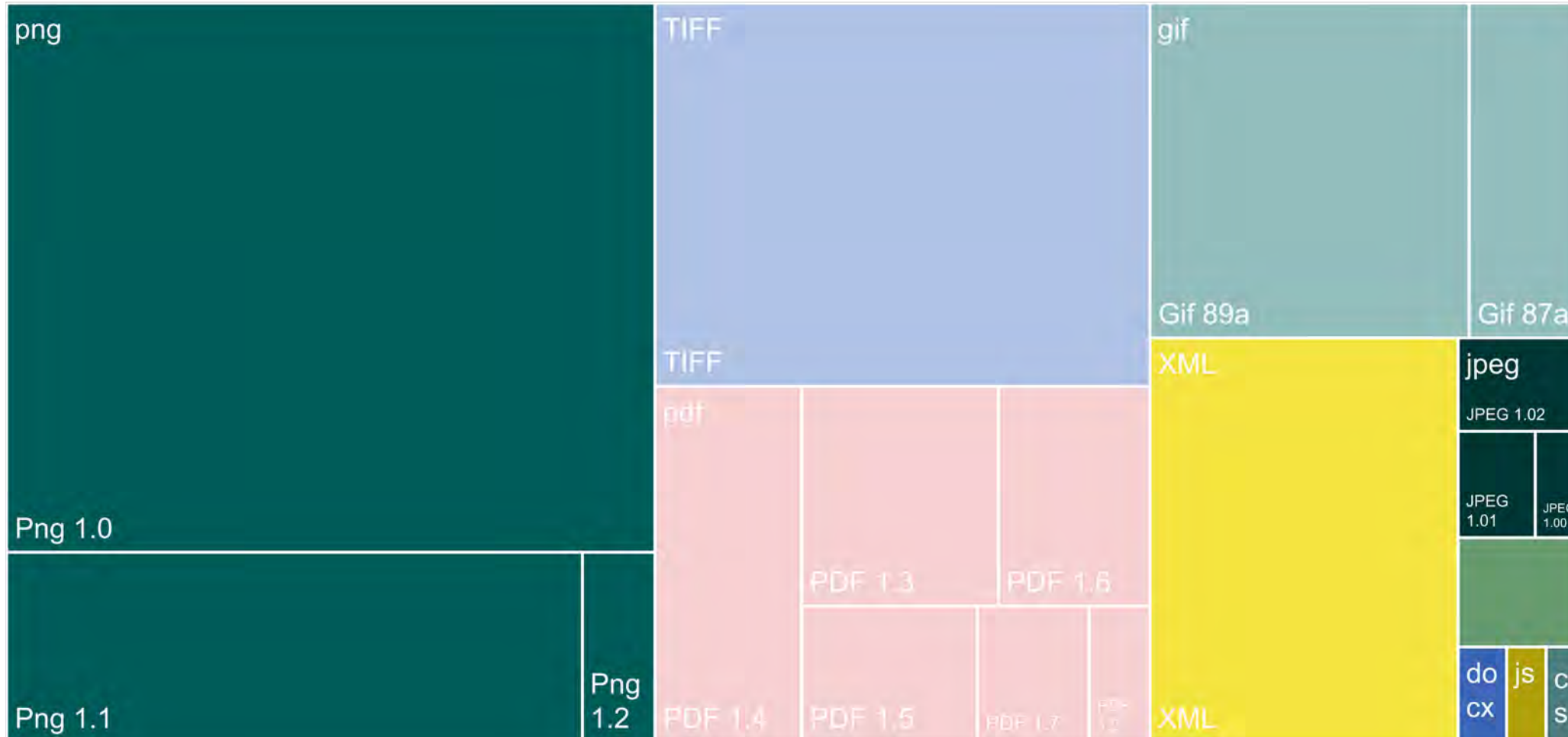
Dauer Entpacken der 290.114 zip-Archive: 5 Tage

Entpackte Daten:

5.2 Millionen Artikel (2.391 Journals) mit
51.8 Millionen Dateien
ca. 11 TB



Top 20 Gruppierung nach Formatfamilien



20 verschiedene Dateiformate decken 91% der Dateien ab

Dateiformate werden **versionsgenau** bestimmt

Dateiformate bilden Grundlage für Archivierungsmaßnahmen

Long Tail der Dateiformate im TIB Digitalen Langzeitarchiv

Insgesamt 254 verschiedene Dateiformate

**44 Dateiformate mit
> 1.000 Dateien,**

z.B.

xslx (6372)

avi (4502)

pptx (1982)

**23 Dateiformate mit
100 - 999 Dateien,**

z.B.

html < 2.0 (828)

mp3 (566)

csv (482)

**69 Dateiformate mit
10 - 99 Dateien, z.B.**

Flow Cytometry Standard File (83)

Crystallographic Information Framework (67)

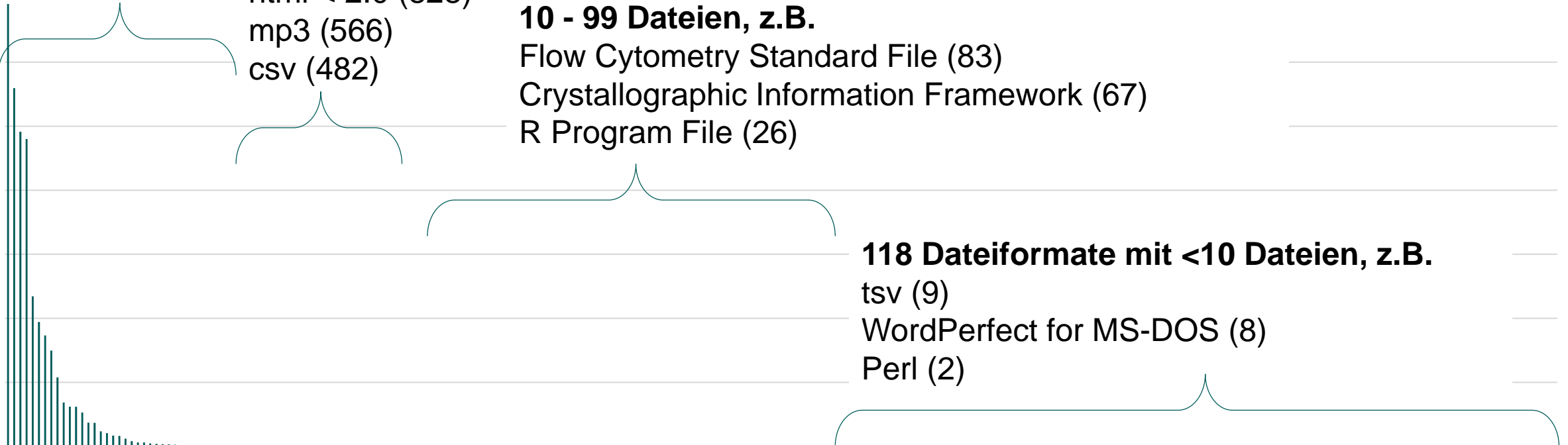
R Program File (26)

118 Dateiformate mit <10 Dateien, z.B.

tsv (9)

WordPerfect for MS-DOS (8)

Perl (2)

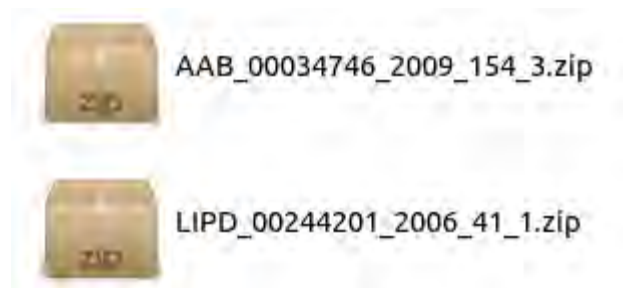


Vollständigkeitsabgleich – Inventory-File als Input

Datenlieferung

(Ist-Zustand)

- 1 ZIP = 1 Issue
- 1 Artikel = 1 Unterordner
- Dateiname nach Schema:
JGC_ISSN_DATE_VOL_ISSUE
- Jeder Artikel enthält XML mit Metadaten
- in den meisten Fällen auch PDF
- in vielen Fällen Unterordner (Anhänge, Ressourcen)

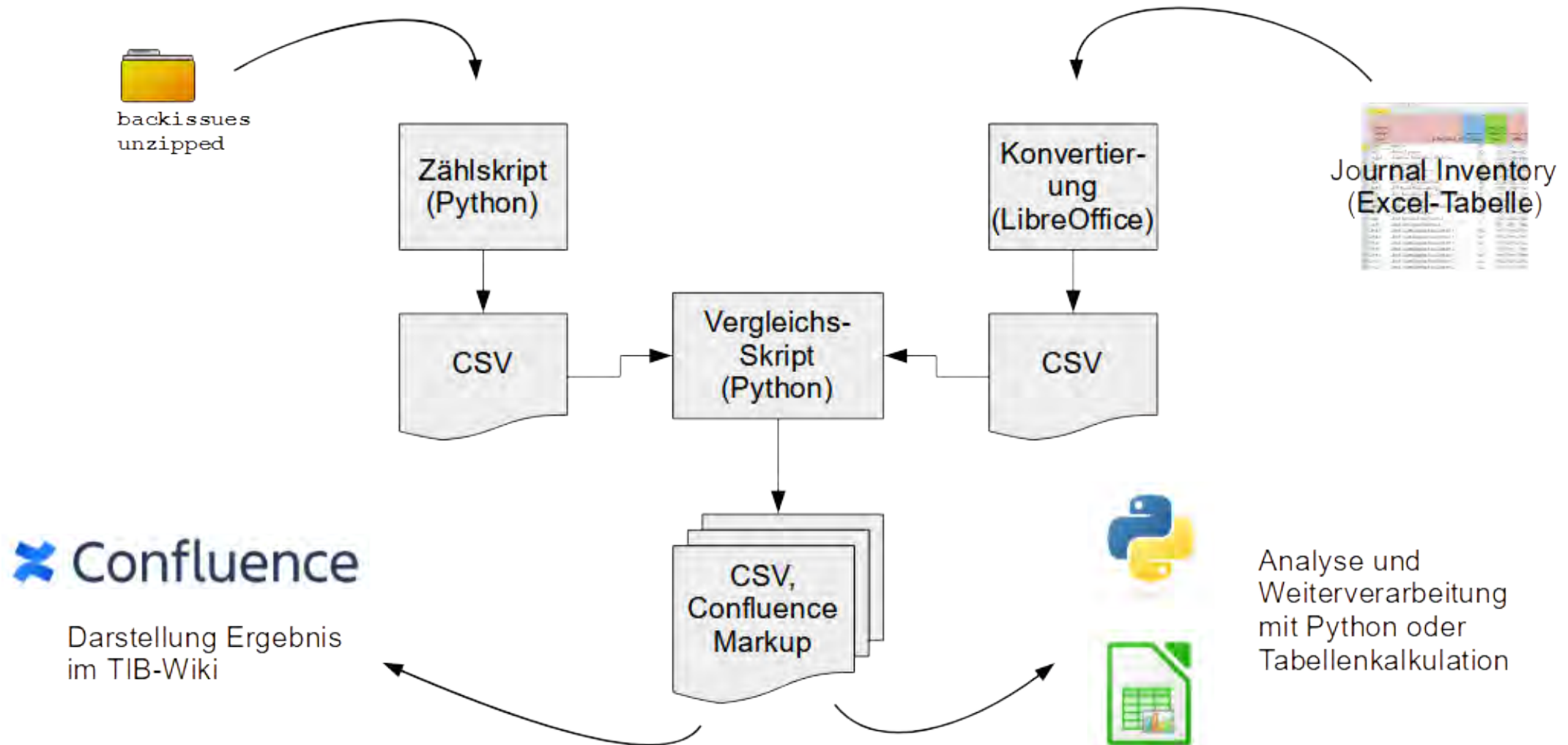


Elemente im Wiley Journal Inventory

(Informationen zum Soll-Zustand)

- Journal group code (JGC)
- publication_title
- print_identifier (ISSN)
- date_first_issue_online
- num_first_vol_online
- num_first_issue_online
- date_last_issue_online
- num_last_vol_online
- num_last_issue_online
- number of issues
- number of articles
- number of PDF
- number without PDF

Ist auch alles da? Vollständigkeitsabgleich – „Global“

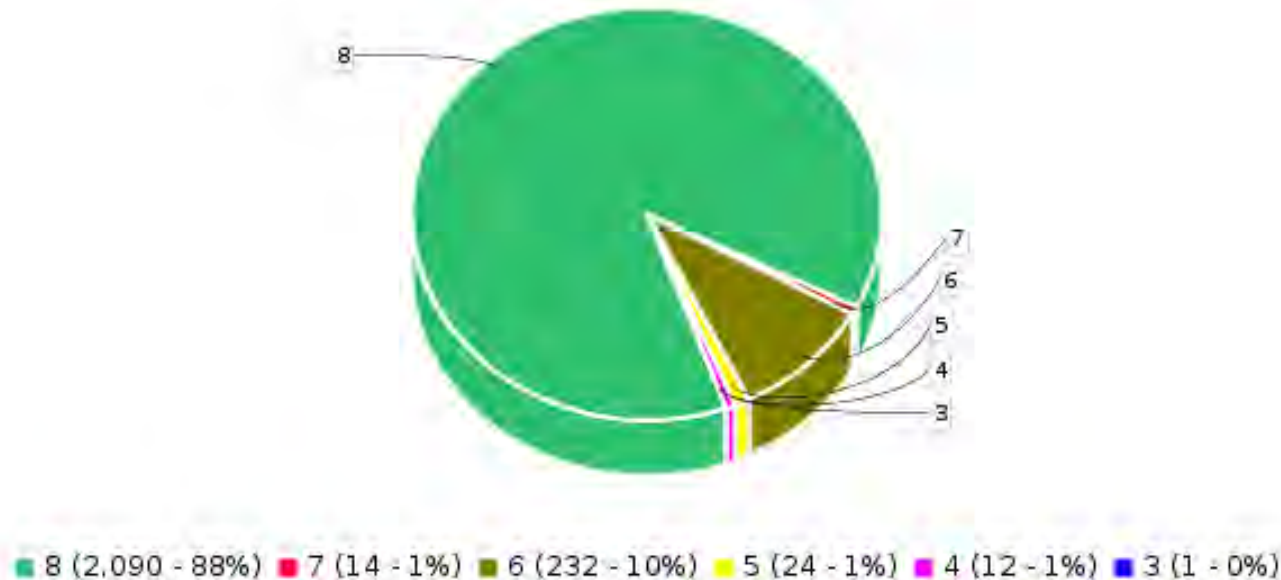


Vollständigkeitsabgleich – „Global“ Ergebnis Wiley Back Issues

Match Level

Der Match Level ist ein Zähler, der bei einem erfolgreich abgeglichenen Wert hochgezählt wurde. Er ist ein Indikator für den Grad der Übereinstimmung eines Journals. Da nur je 8 Werte verglichen wurden bedeutet ein Match-Level von 8 die maximale mögliche Übereinstimmung innerhalb dieses Abgleichs. Abweichungen jenseits des des Abgleichs – zum Beispiel bei den Issues – sind theoretisch möglich und müssen beim Pre-Ingest geprüft werden.

Verteilung Journals pro Match Level



2373 Zeitschriften berücksichtigt
(Screenshot aus internem Wiki)

Abgeglichene Daten

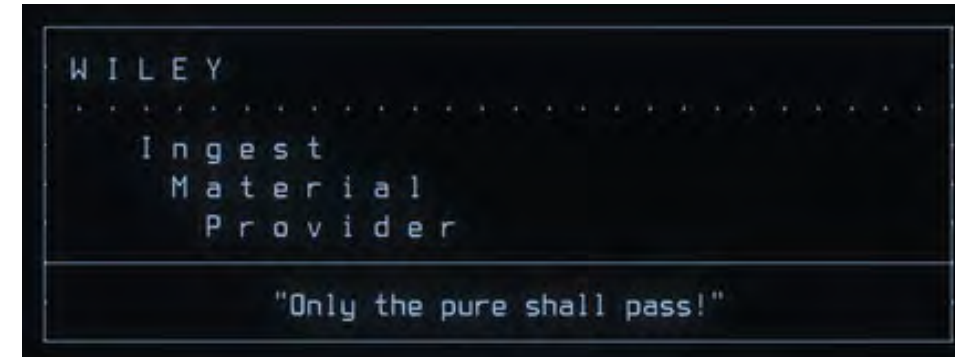
- Journal Group Code
- Print ISSN
- Date First Issue
- Date Last Issue
- Number of Issues
- Number of Articles
- Number of Articles with PDF-Representation
- Number of Articles without PDF-Representation

= 8 Elemente

Ist auch alles da? Vollständigkeitsabgleich – „Detailliert“

Tool „Wiley IMP“ Funktionen

- Abgleich verschiedener (Meta-)Daten untereinander
- Kontrolle der Ordnerstruktur
- Kontrolle des obligatorischen Inhalts (XML/PDF)
- Warnung bei leeren oder sehr kleinen Dateien
- Anzahl der Dateien, Größe des Journals (MiB)



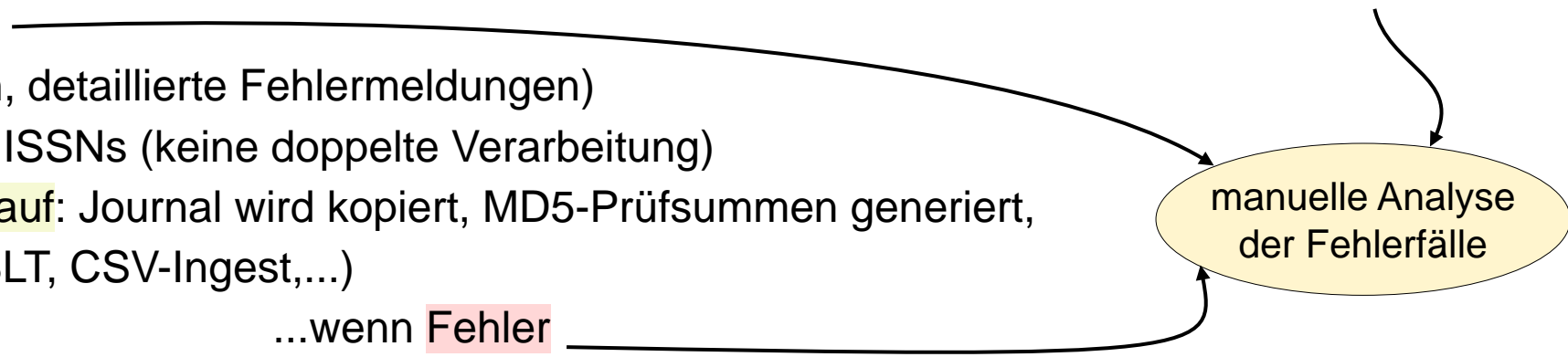
Ergebnisse

- Report-File (Überblick)
- Log-File (chronologisch, detaillierte Fehlermeldungen)
- merkt sich verarbeitete ISSNs (keine doppelte Verarbeitung)
- Bei fehlerfreiem Durchlauf: Journal wird kopiert, MD5-Prüfsummen generiert, Weiterverarbeitung (XSLT, CSV-Ingest,...)
...wenn Fehler

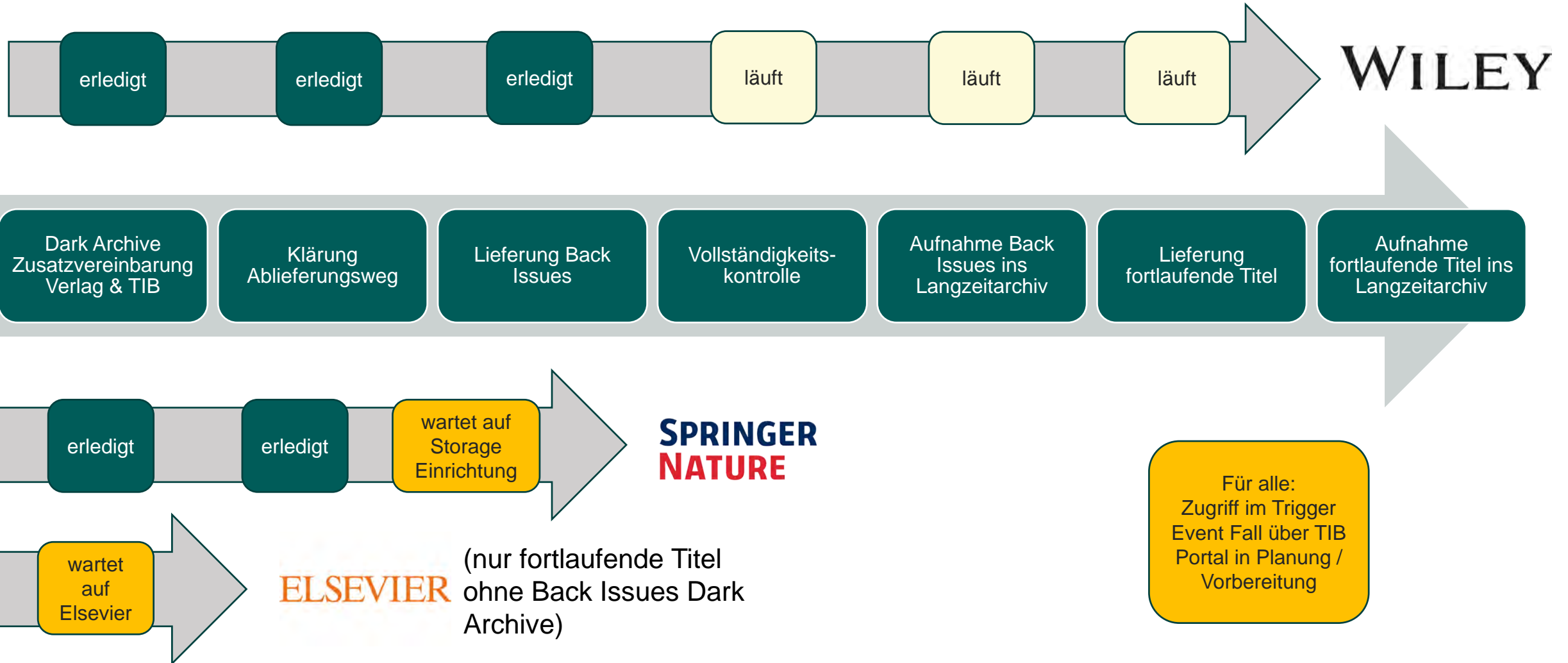


backissues
unzipped

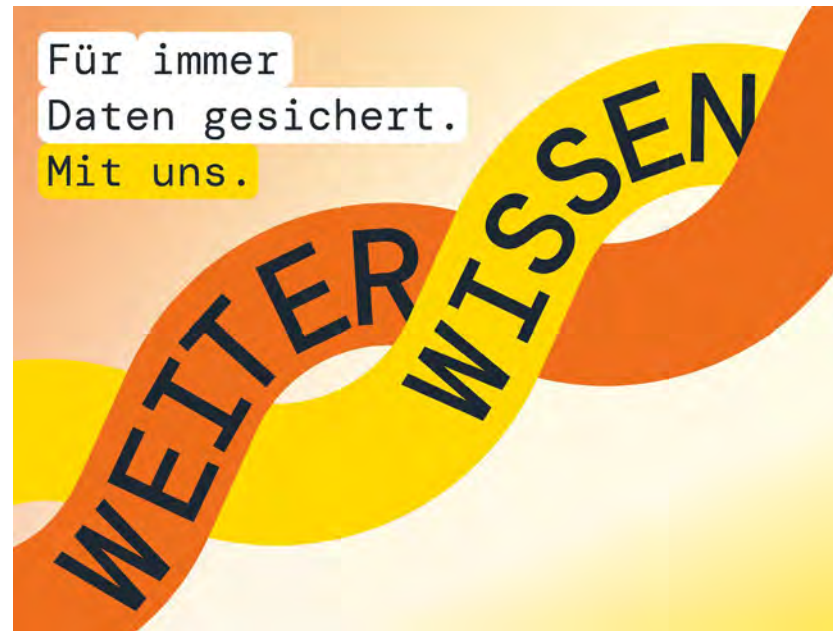
manuelle Analyse
der Fehlerfälle



IST-Stand DEAL Dark Archives an der TIB



Vielen Dank! Fragen?



Kontakt

Micky Lindlar (they/them)

 micky.lindlar@tib.eu

 @mickylindlar@digipres.club



Creative Commons Namensnennung 3.0 Deutschland
<http://creativecommons.org/licenses/by/3.0/de>