

Neue Wege zur FAIRen Archivierung von digitalen Forschungsdaten aus Bayern

Tanja Weibulat



Kooperation zweier NFDI4Biodiversity-Partner



Technische Kooperation zweier Einrichtungen
unter dem gleichen Bayerischen Staatsministerium



- Fachbehörde für das Archivwesen im Freistaat Bayern
- Aufgabe des digitalen Archivs:
Archivierung der Informationen von bleibendem Wert,
d.h. aller entbehrlicher elektronischer Informationen der
Behörden und Gerichte im Freistaat Bayern, die aufgrund
ihrer
 - Rechtserheblichkeit
 - Bedeutung für Wissenschaft und Forschung
 - Bedeutung für Rückgriffe von Justiz und Verwaltung
zeitlich unbefristet aufzubewahren sind.

- Wissenschaftliche
Fachbehörde aus 10
Naturkundemuseen, dem
Botanischen Garten
München und 7
naturhistorischen
Sammlungen
- Fachrichtungen: Zoologie, Geologie,
Paläontologie, Mineralogie, Anthropologie
Paläoanatomie, Botanik und Mykologie
- Eines der größten Herbarien weltweit, genannt
Botanische Staatssammlung München

Artenvielfalt in Herbarien - Mehltaupilzsammlung



Archivierung von Datenbeständen in relationalen Datenbanken anhand des SIARD-Standards





Was ist das SIARD-Format?

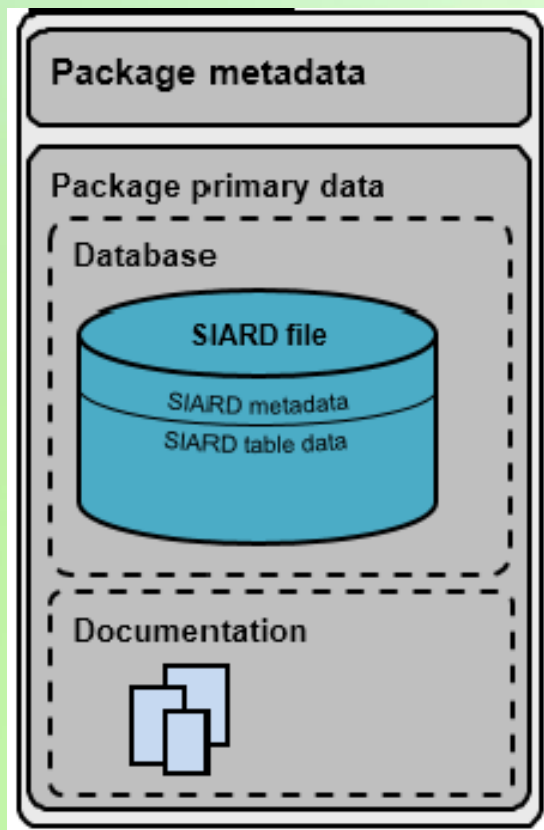
Archivtaugliche Dateiformate

• Textdaten (unstrukturiert)	TEXT
• (Text-)Daten (strukturiert)	XML
• Druckbare Dokumente	PDF/A
• Tabellarische Daten	CSV
• Relationale Datenbanken	SIARD
• Rasterbilder	JPEG2000, TIFF
• Audiodaten	WAVE
• Videodaten	FFV1, MPEG/4

Merkblatt des Schweizerischen Bundesarchivs (BAR) „Archivtaugliche Dateiformate – Standards für die Archivierung digitaler Unterlagen“ (Stand April 2020): https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf

SIARD-Datenpaket und Abgabestruktur Mehltaupilze

<https://siard.dilcis.eu/SIARD%202.2/SIARD%202.2.pdf>



Abgabestruktur AIP BSMeryscoll¶

1-ZIP-Datei-PROLOG¶

****Dateiname: "BAYHSTA_PROLOG_UUID.zip"¶

- 1-x-PDF/A mit Handbuch der entsprechenden DWB-Version¶

****Dateiname: "Handbuch_DWB_[Versionsnr.].pdf"¶

Anmerkung: Im Use-Case für DWB-Collection und DWB-Projects¶

- 1-x-PDF/A DWB-Wikiseite¶

****Dateiname: "Wiki_DWB_[lfd. Nr.][Versionsnr.].pdf"¶

Anmerkung: Im Use-Case für DWB-Collection und DWB-Projects¶

- 1-x-PDF/A mit inhaltlichen Infos zu den Daten der Datenbank (z.B. Bedeutung, Laufzeit usw.)¶

****Dateiname: "Dokumentation_lfd.Nr._Daten_[UUID-der-ZIP-Datei-DATEN].pdf"¶

- 1-CSV mit Datenbankaufbau¶

****Dateiname: "SIARD_Model_Collection_[UUID-der-ZIP-Datei-DATEN].csv"¶

- 1-CSV mit Datenbankaufbau¶

****Dateiname: "SIARD_Model_Projects_[UUID-der-ZIP-Datei-DATEN].csv"¶

- 1-PDF/A mit Datenbankstruktur¶

****Dateiname: "DWB_Model_Collection_[UUID-der-ZIP-Datei-DATEN].pdf"¶

- 1-PDF/A mit Datenbankstruktur¶

****Dateiname: "DWB_Model_Projects_[UUID-der-ZIP-Datei-DATEN].pdf"¶

1-ZIP-Datei-DATEN¶

****Dateiname: "BAYHSTA_DATEN_UUID.zip"¶

- 1-SIARD-Datei der Daten¶

****Dateiname: "DWB_[Bezeichnung-der-DB]_[Bezeichnung-der-Collection].SIARD"¶

- 1-SIARD-Datei der Daten¶

****Dateiname: "DWB_[Bezeichnung-der-DB]_[Bezeichnung-der-Projects].SIARD"¶

- 1-CSV-Datei mit grundsätzlichen Erschließungsmetadaten zum Gesamtpaket (ZIP-Datei-PROLOG)¶

****Dateiname: "BAYHSTA_Erschließung_UUID.csv"¶

- Datenhaltende Stelle¶
- Entstehungsjahr-Beginn (Zeitpunkt Anlage des ersten Datensatzes)¶
- Entstehungsjahr-Ende (Zeitpunkt Datenschnitt)¶
- Titel¶
- Allgemeine Beschreibung¶
- UUID¶
- Lizenz¶
- Weitere Nutzungseinschränkungen¶
- Herkunft der Metadaten¶

¶

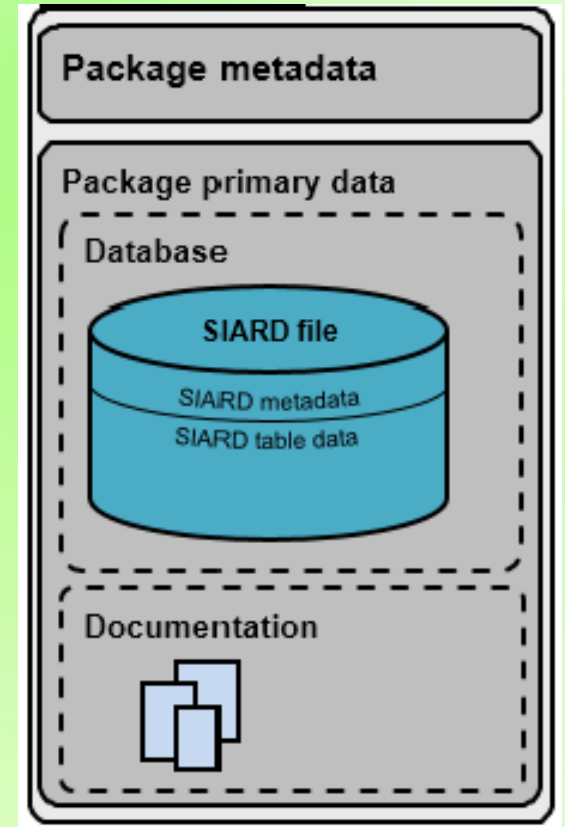
Archival
Information
Package

Ziel der SIARD-Archivierung

→ datenbank-unabhängige **logische, semantische und relationale** Archivierung zur Verfügbarkeit "**auf ewig**"

1. Betrifft die **kompletten (!) Daten** eines relationalen Datenbestandes inkl. History, personenbezogene und andere sensible, nicht zur Veröffentlichung vorgesehene Daten aus DWB-SQL-Datenbanken als Zeitschnitt (Version)

2. Betrifft DWB Entity Relations Diagramme, große Teile der dokumentierten DWB-Datenbank-Logik, weitere **Dokumentationen** zur DWB (Manual) und inhaltliche Beschreibung des Datenbestandes („About“)

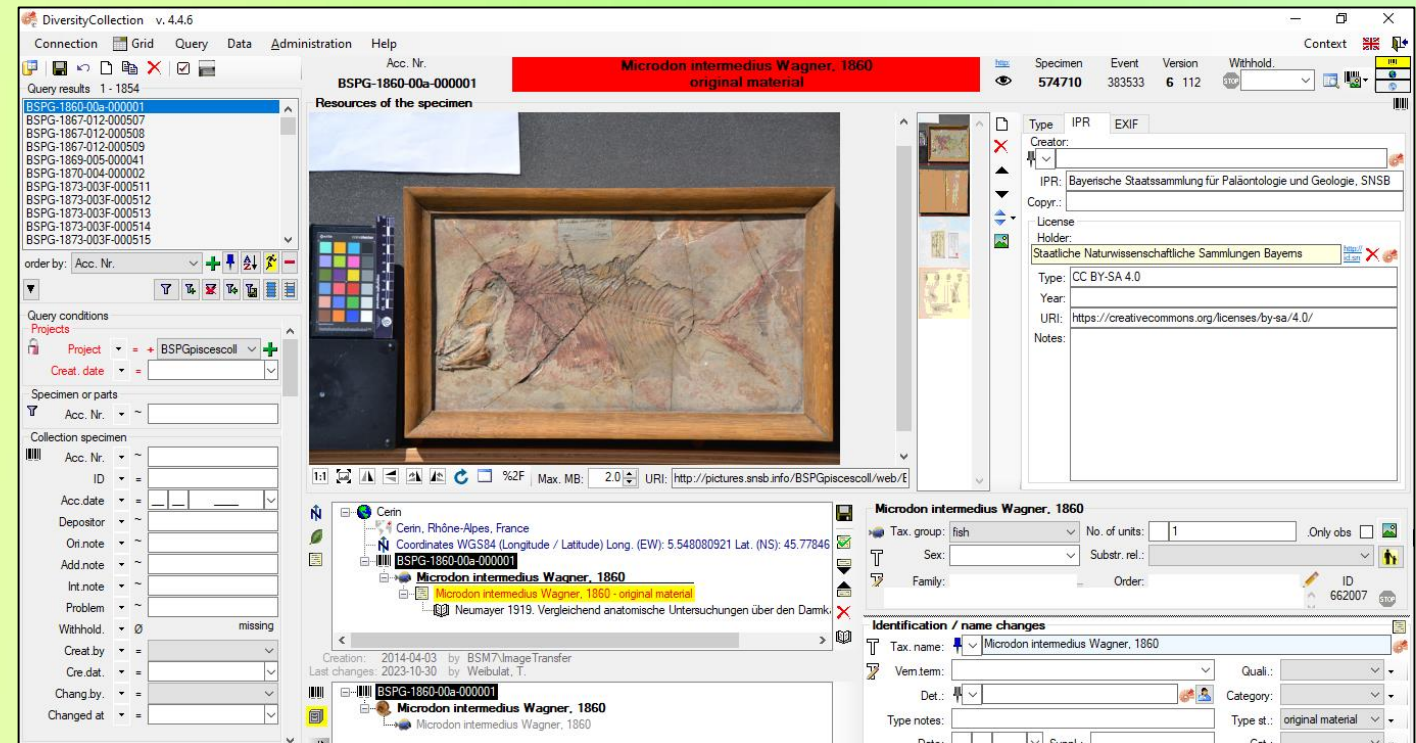


<https://siard.dilcis.eu/SIARD%202.2/SIARD%202.2.pdf>

Fortsetzung der technischen Kooperation

SIARD-Archivierung der kompletten DWB-Daten zu zwei weiteren SNSB-Sammlungsdatenbeständen:

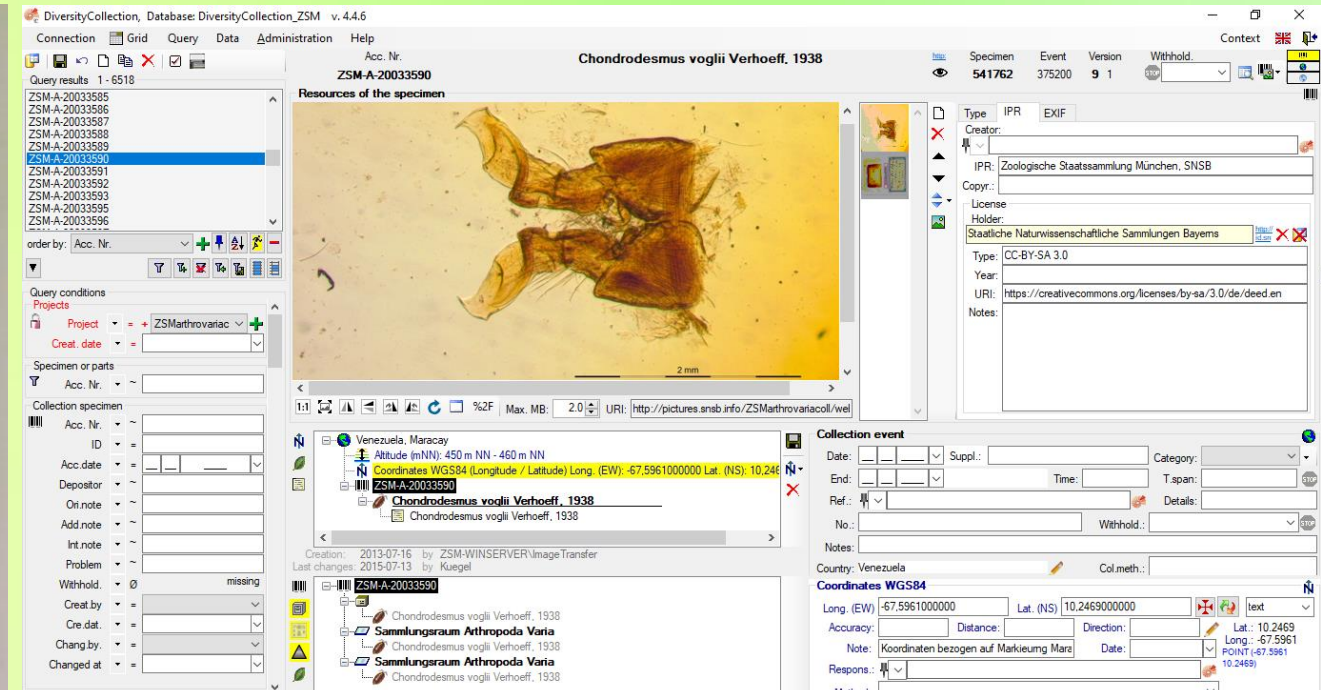
- Fossile Fische der SNSB – Bayerischen Staatssammlung für Paläontologie und Geologie



Fortsetzung der technischen Kooperation

SIARD-Archivierung der kompletten DWB-Daten zu zwei weiteren SNSB-Sammlungsdatenbeständen:

- Sammlung der Gliederfüßer der SNSB – Zoologischen Staatssammlung München
 - Mikroskopische Objektträger der Mundwerkzeuge von Milben



Ausblick



NFDI 4
BIODIVERSITY



- Technische Kooperation zur SIARD-Archivierung in Phase 2 von NFDI4Biodiversity: weitere Datenpakete aus den SNSB
- Kooperation mit weiteren Archiven und fachspezifische Datenzentren
- Konsortium-übergreifendes SIARD-Training im Rahmen von NFDI
- SIARD-Archivierung von anderen Typen von Forschungsdaten und anderen relationalen Datenbanken
- Ergebnisse werden publiziert

Beteiligte Personen

- Generaldirektion der Staatlichen Archive Bayerns:



- Giada Matheisen
- Michael Puchta
- Markus Schmalzl
- Anna Lisa Schwartz

Danke für Ihre Aufmerksamkeit!

- Staatliche Naturwissenschaftliche Sammlungen Bayerns:



- Stefan Seifert
- Dagmar Triebel
- Tanja Weibulat
- Markus Weiss