DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Data Preparation Tool neo

Datenaufbereitungssoftware für Archivdaten

KIM Workshop 2019
02.04.2019, Mannheim

Data Preparation Tool

Motivation

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- große Anzahl archivischer Datenpartner mit heterogener Datenhaltung und Erschließungspraxis
- Wunsch nach regelmäßigen Updates in DDB und Archivportal-D
- Neue Anforderungen an Datenlieferungen, die nicht immer erfüllt werden können
    - vor allem: fehlende Schnittstellen in älterer Erschließungssoftware

Data Preparation Tool

Zielsetzung

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Dezentralisierung der Datenaufbereitung
    - insbesondere Update-Lieferungen durch Datengeber selbst aufbereiten lassen
- Workflows in der Fachstellenarbeit zentral abbilden
- Nachnutzbarkeit erhöhen
    - Nachnutzung providerspezifischer Anpassungen durch stärkere Modularisierung
    - Einfache Einbindung neuer Ein-/Ausgangsformate

- Ziel: Archive können Updates selbständig aufbereiten
    - → mehr Kontrolle über Ihre Daten
    - → **Ergänzung** zur individuellen Unterstützung, *kein* Ersatz!

Data Preparation Tool

Besonderheiten von Archivdaten

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

050

Beschluß
der Volkskammer der
Deutschen Demokratischen Republik
über den Beitritt der Deutschen Demokratischen Republik
zum Geltungsbereich des Grundgesetzes der
Bundesrepublik Deutschland
vom 23. August 1990

Die Volkskammer erklärt den Beitritt der Deutschen Demokratischen Republik zum Geltungsbereich des Grundgesetzes der Bundesrepublik Deutschland gemäß Artikel 23 des Grundgesetzes mit Wirkung vom 3. Oktober 1990.

Sie geht dabei davon aus,

- daß die Beratungen zum Einigungsvertrag zu diesem Termin abgeschlossen sind,

- die 2 + 4 - Verhandlungen einen Stand erreicht haben, der die außen- und sicherheitspolitischen Bedingungen der deutschen Einheit regelt,

- die Länderbildung soweit vorbereitet ist, daß die Wahl in den Länderparlamenten am 14. Oktober 1990 durchgeführt werden kann.

Vorstehender Beschluß wurde von der Volkskammer der Deutschen Demokratischen Republik in ihrer 30. Tagung am 23. August 1990 gefaßt.

Berlin, 23. August 1990

Die Präsidentin der Volkskammer
der Deutschen Demokratischen Republik

Nach emotionsgeladenen Diskussionen und intensiven Beratungen stimmt die Mehrheit der Abgeordneten der Volkskammer in ihrer 30. Tagung am 23. August 1990 für den Beitritt zur Bundesrepublik nach Artikel 23 des Grundgesetzes. Bedingung ist, dass die notwendigen Voraussetzungen dafür geschaffen werden, wie zum Beispiel die Gründung der Länder oder der erfolgreiche Abschluss der Zwei-plus-vier-Verhandlungen.

Beispiel für ein archivisches Einzelobjekt: Akte

Data Preparation Tool

Besonderheiten von Archivdaten

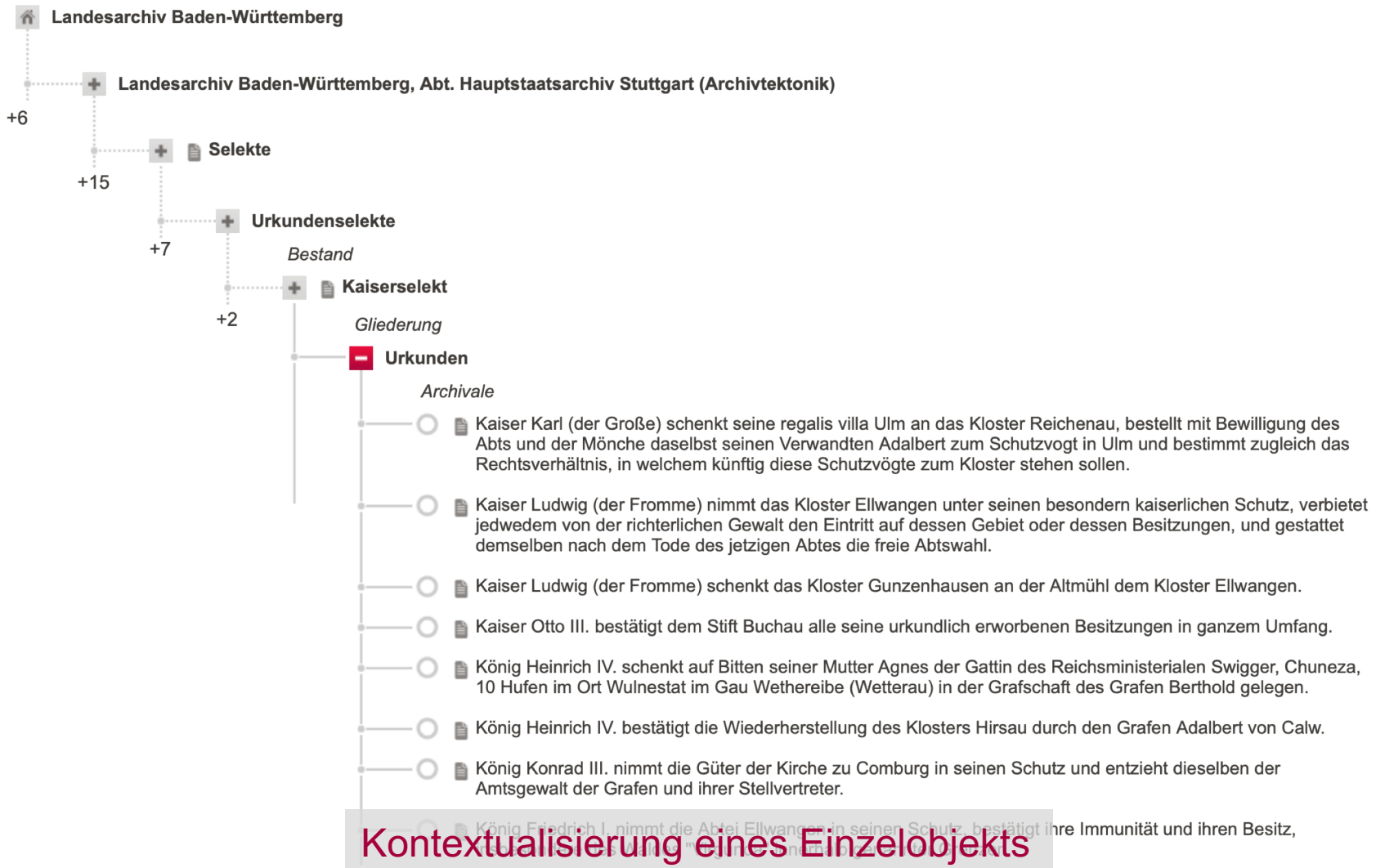
DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Beispiel für ein archivisches Einzelobjekt: Urkunde

Data Preparation Tool

Besonderheiten von Archivdaten

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Kontextualisierung eines Einzelobjekts durch übergeordnete Hierarchien

Data Preparation Tool

Besonderheiten von Archivdaten

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Administrative Informationen
- Bestandsbezogene Informationen
  - Gliederungsebene 1
    - Gliederungsebene 1.1
      - Verzeichnungseinheit 1

Data Preparation Tool

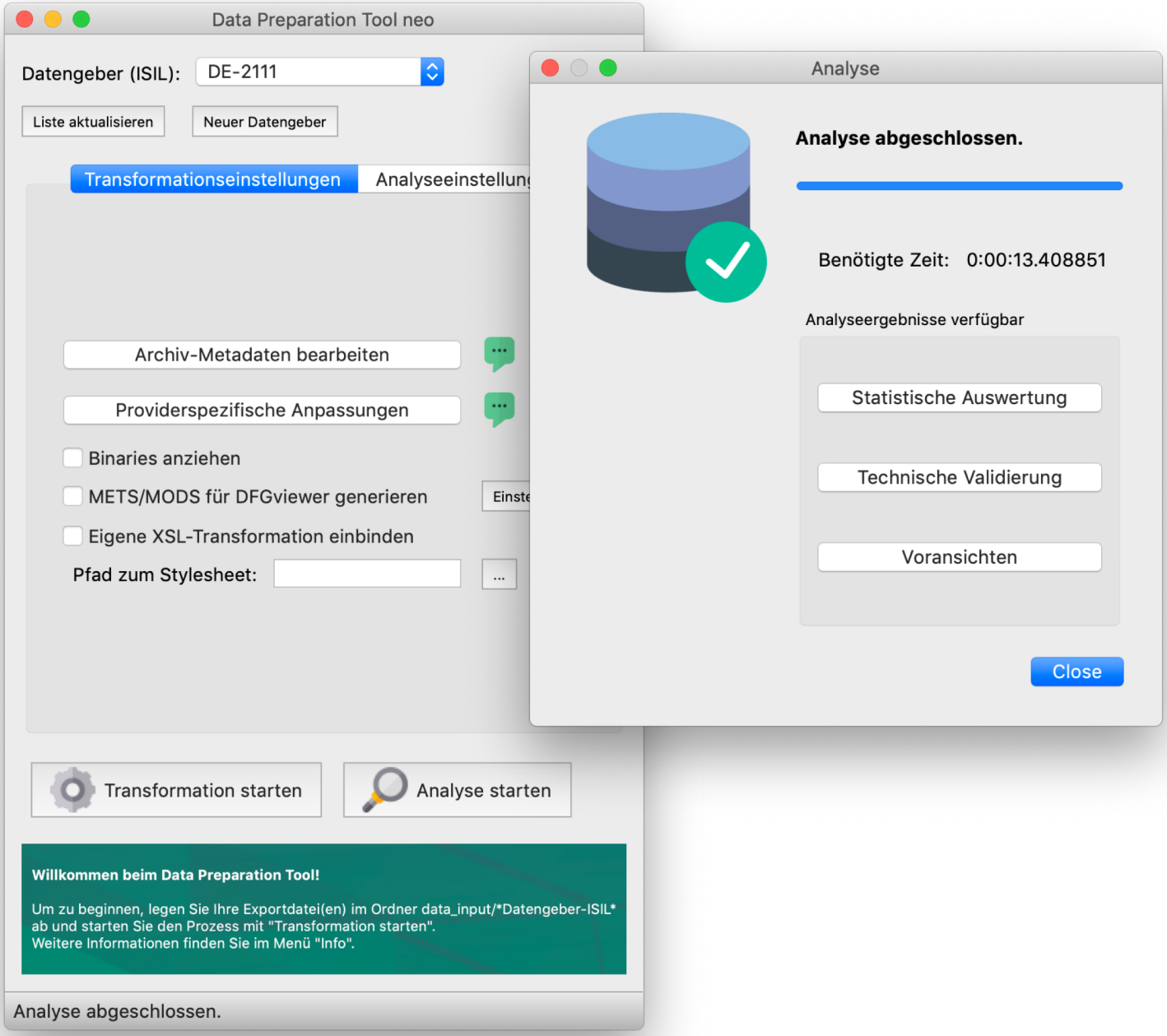
Konzept und Funktionalität

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- **Transformation** beliebiger Quelldaten (vorzugsweise XML) nach EAD(DDB)
- **Vorprozessierung** für den DDB-Ingest
  - Anziehen der Binaries, OAI-Harvesting, Normalisierung, METS/MODS-Erzeugung zum Bedienen des DFG-Viewers
- **Plausibilitätsprüfung** der Datenlieferung
  - Identifier, Hierarchieverknüpfung, obsolete Objekte
- Auszeichnung der **Rechte-** und **Aggregatorinformation**

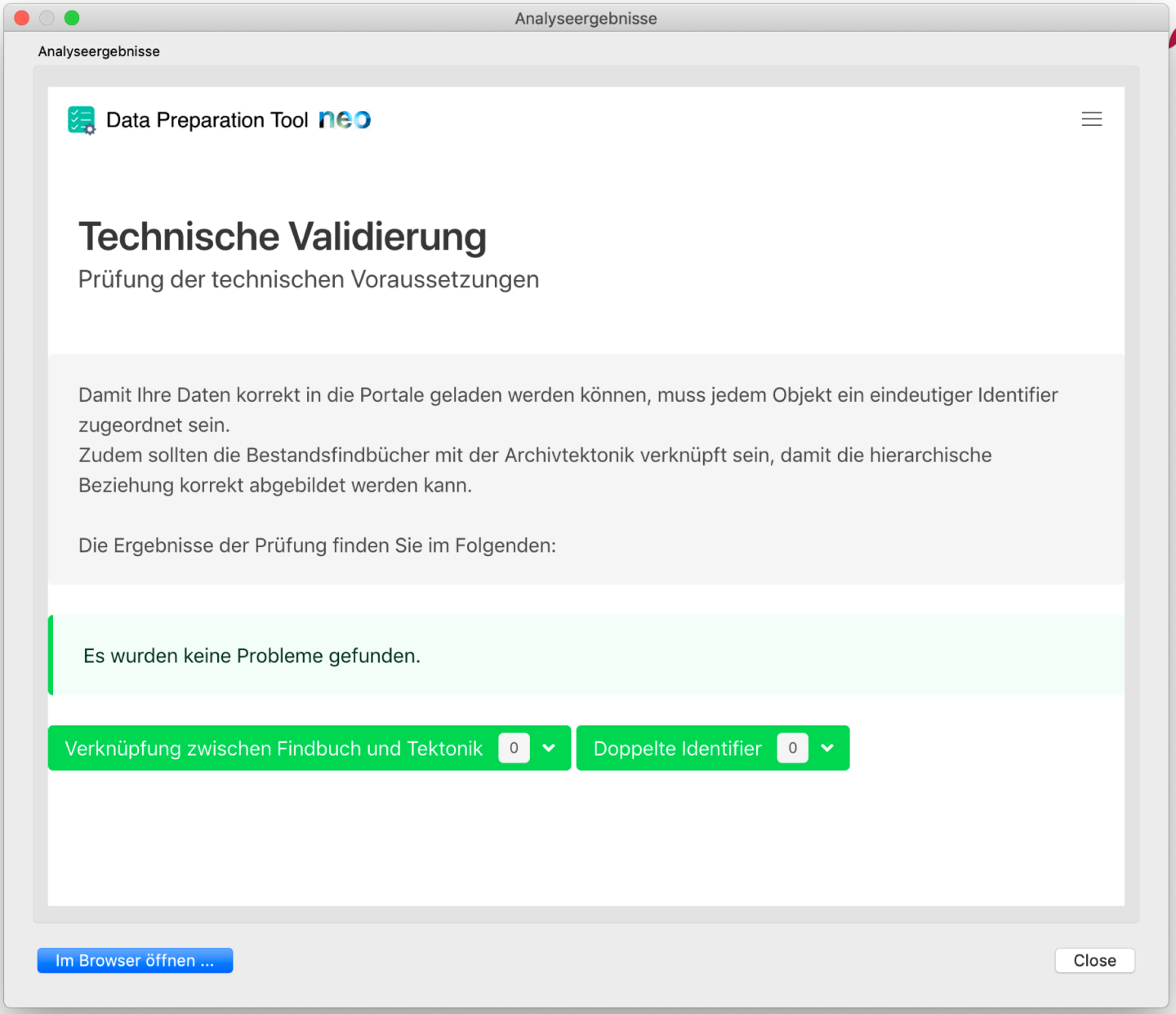
Data Preparation Tool

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Data Preparation Tool

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Data Preparation Tool

Konzept und Funktionalität

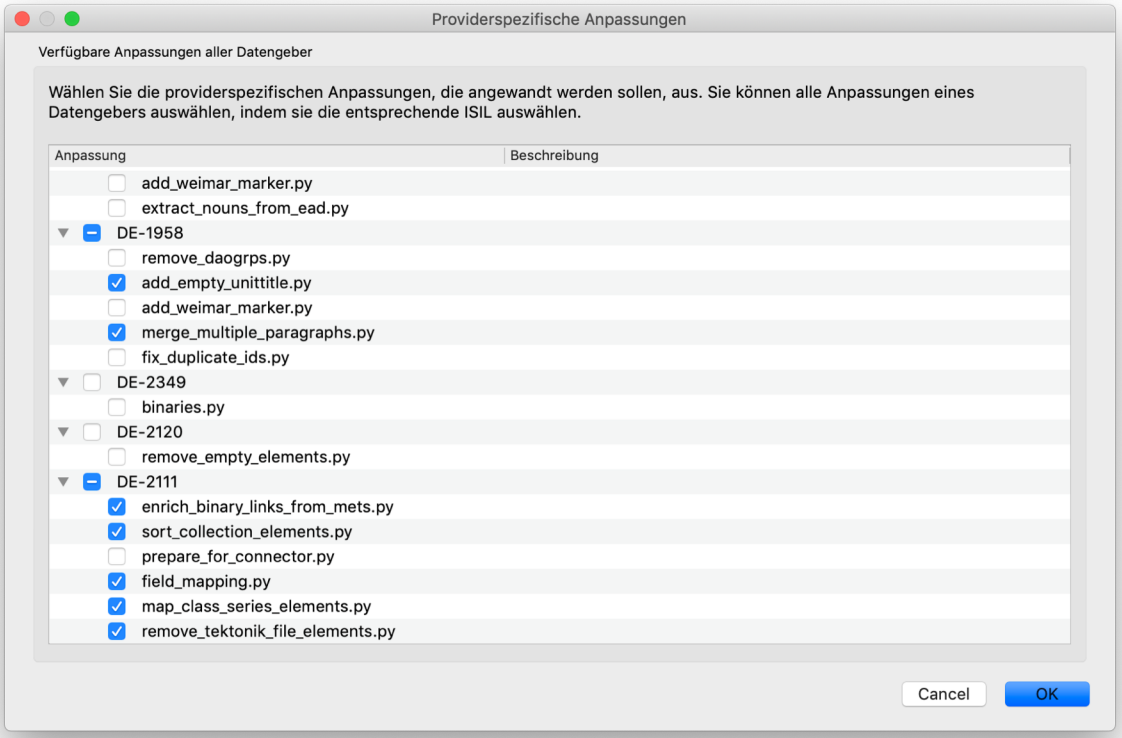
DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

– Providerspezifische Anpassungen

feingranulare, nachnutzbare Mappinganpassungen

kombinierbar über mehrere Datengeber hinweg

Python-Skripte bieten große Flexibilität

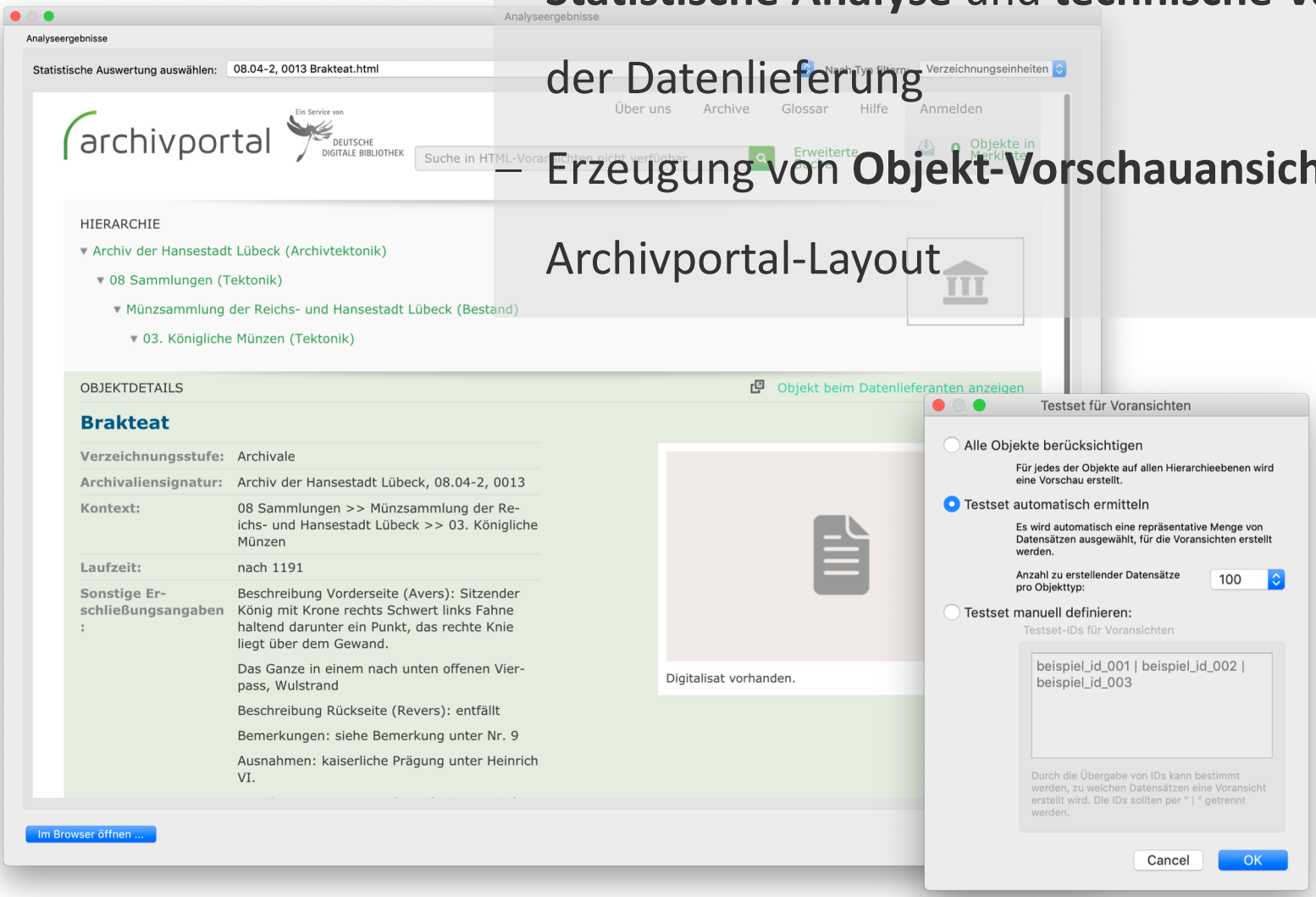


Data Preparation Tool

Konzept und Funktionalität

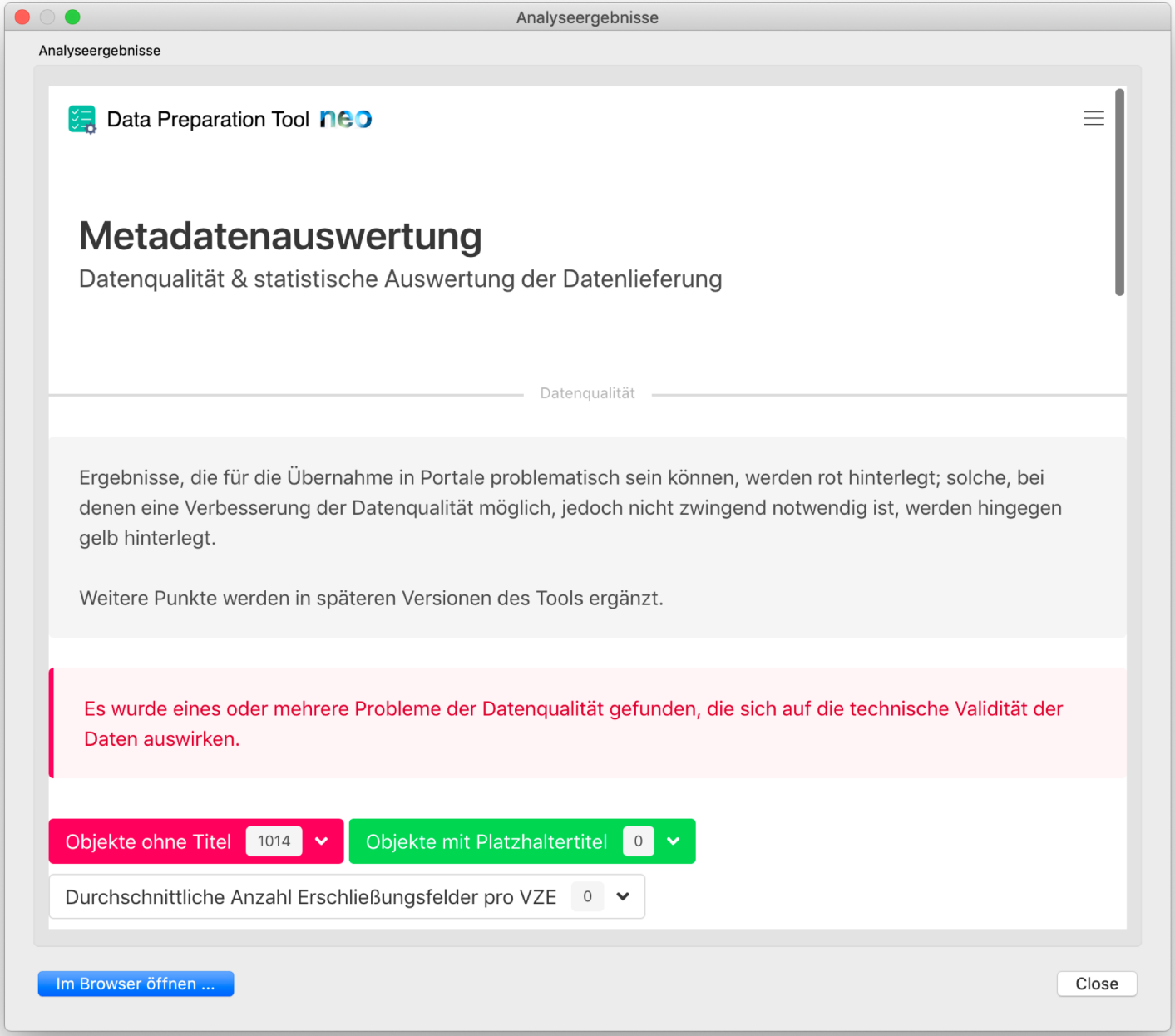
DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- **Statistische Analyse** und **technische Validierung** der Datenlieferung
- Erzeugung von **Objekt-Vorschauansichten** im Archivportal-Layout



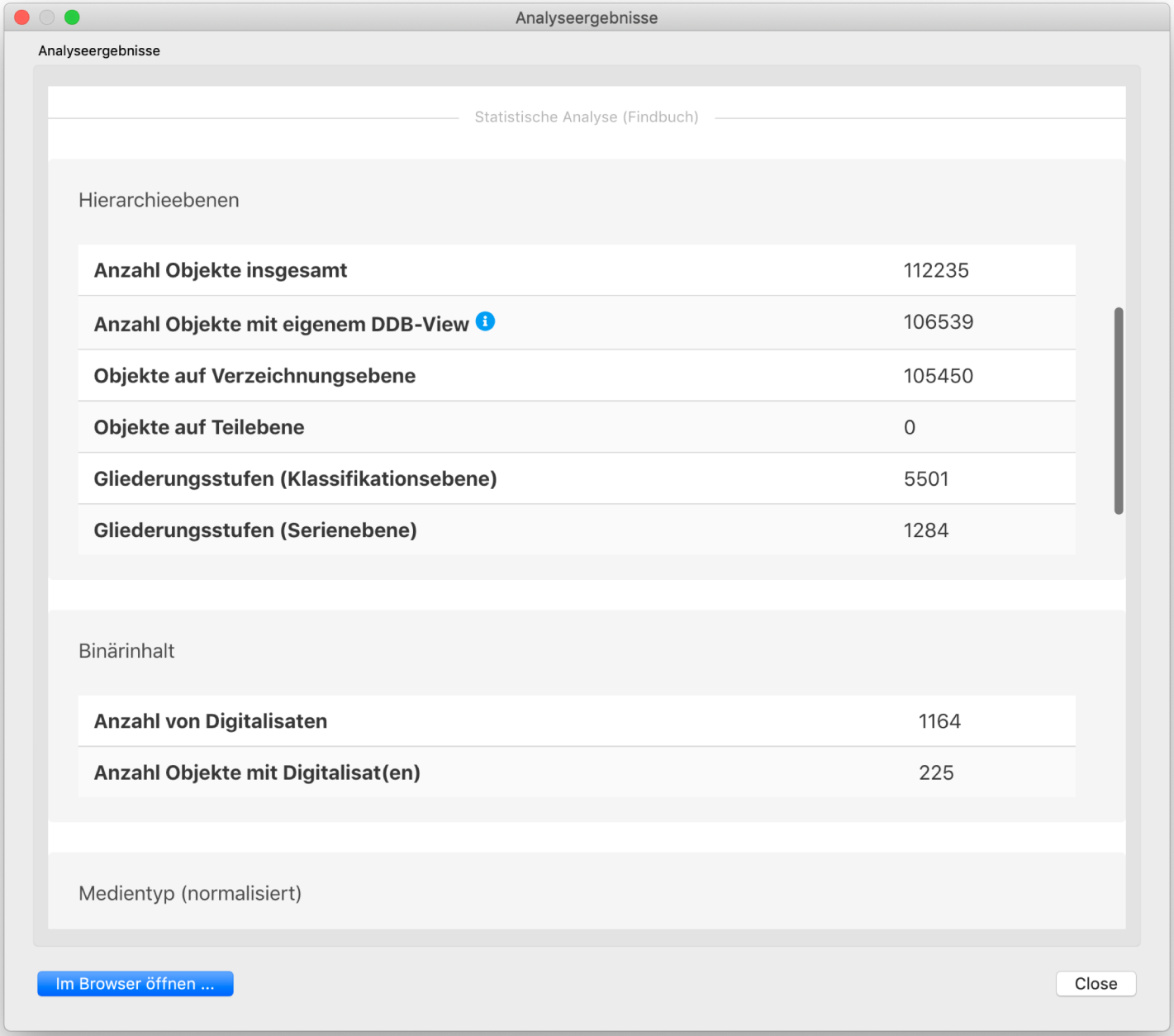
Data Preparation Tool

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



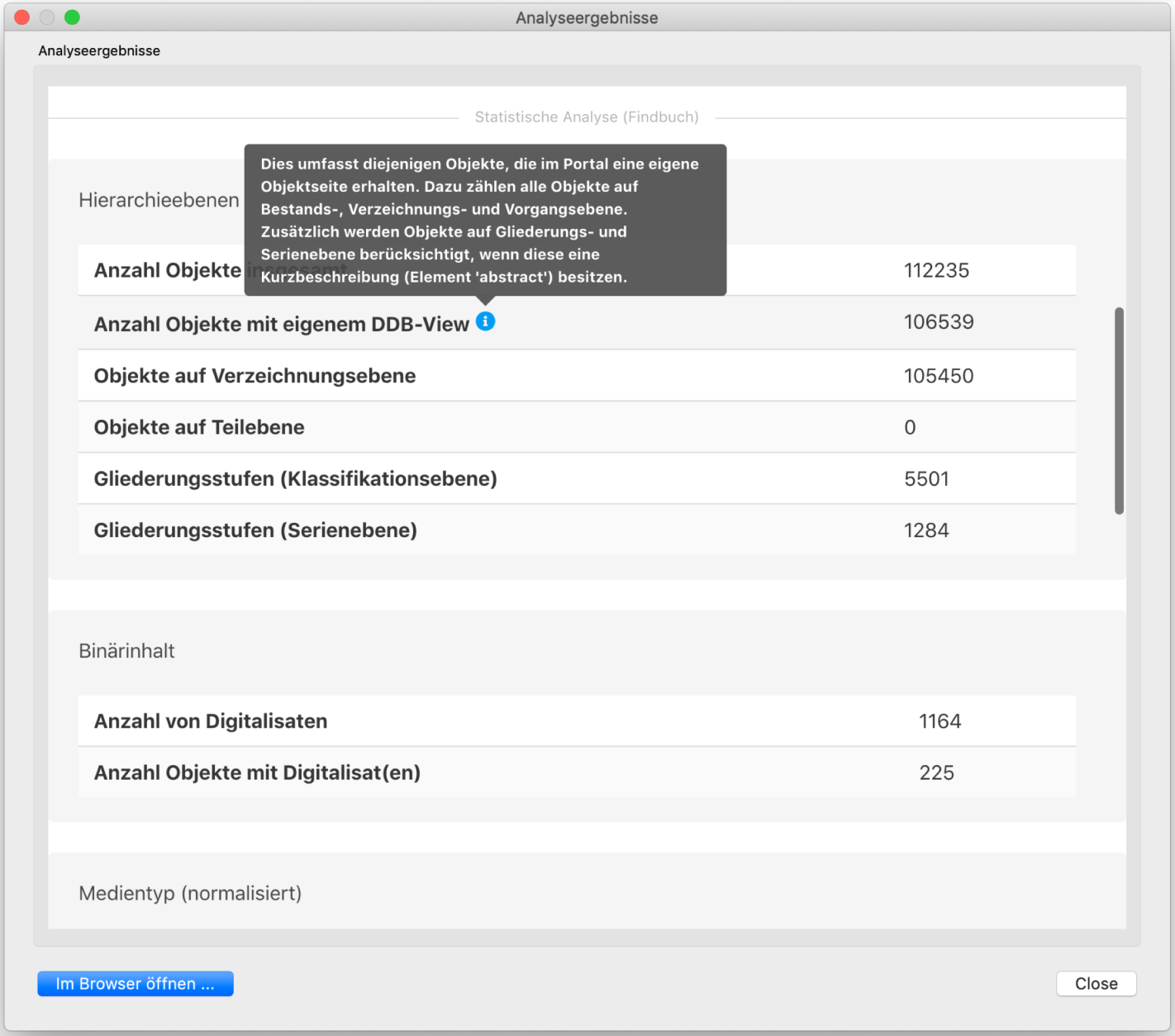
Data Preparation Tool

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Data Preparation Tool

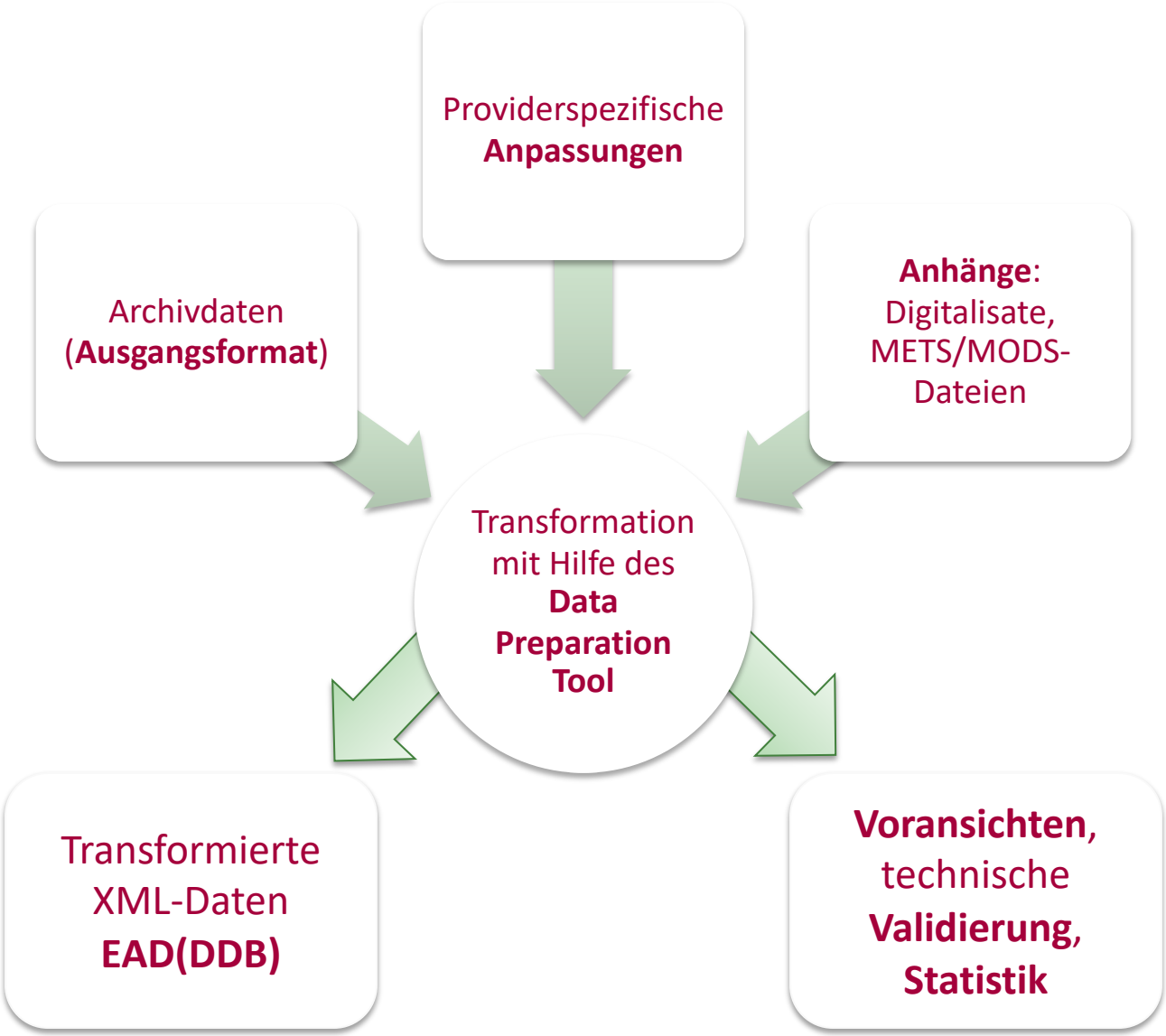
DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Data Preparation Tool

Konzept und Funktionalität

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



Data Preparation Tool

Open-Source-Technologien

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

python™

lxml

Requests
http for humans

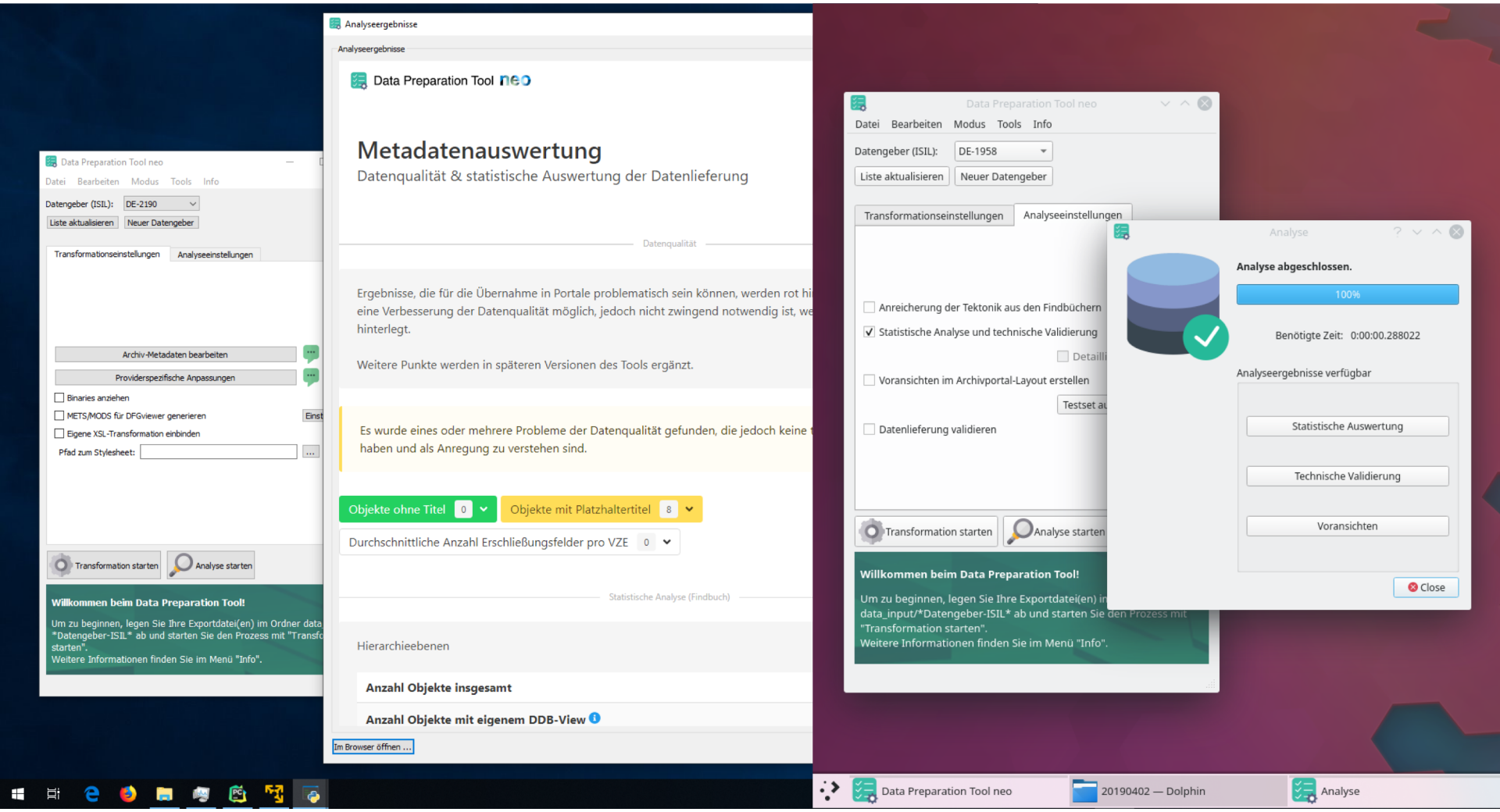
BULMA.io

Qt

Data Preparation Tool

Cross-Platform

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online



DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Visuelle Validierung

Verbesserung der Persistenz von Objekt-Identifiern

Unterstützung der „großen“ EAD-Version 3.0

Data Preparation Tool

Weiterentwicklung: Visuelle Validierung

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Erweiterung durch eine **visuelle Validierung**
  - analog zu XML-Schema-Validierung, aber …
    - menschenlesbare Fehlermeldungen
    - Gruppierung der Ergebnisse nach Priorität
    - Zusammenfassung der Meldungen zu einer übergreifenden Aussage zur Verwendbarkeit der Datenlieferung
  - Zielgruppe: Fachstelle, Datengeber, Aggregatoren und Softwarehersteller

Data Preparation Tool

Weiterentwicklung: Stabile Identifier

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Problematik der nicht immer gegebenen **stabilen Identifier** im Archivbereich, jedoch Voraussetzung für geplante Projekte
- Umsetzung einer **Objekt-ID-Konkordanz** bereits seit längerem ein Desiderat
- Ziel: Rückführung geänderter IDs auf Ihre ursprüngliche ID, wenn jeweiliges Objekt schon einmal geliefert
- Methode: Gewichteter Abgleich über Feldinhalte und Kontext

Data Preparation Tool

Weiterentwicklung: Stabile Identifier

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

```
<c level="file" id="A91x607657551765441201611081337321 40">
    <did>
        <unitid>1</unitid>
        <unittitle>Abel, Fritz</unittitle>
        <unitdate normal="1948-01-01/1954-12-31">01.01.1948 -
            31.12.1954</unitdate>
        <abstract type="Enthält">Enthält u.a.: Personenstandsunterlagen,
            Arztabrechnungen der Ehefrau</abstract>
        <physdesc><extent>1 cm</extent></physdesc>
    </did>
    <daogrp>
        <daodesc>
            <list>
                <item><name>B2_1</name><genreform>SONSTIGES</genreform></item>
            </list>
        </daodesc>
    <daoloc xlink:role="image_full" xlink:href="bilder/B2_1.pdf"/></daogrp>
</c>
```

Für Hash-Abgleich relevante EAD-Felder
(u.a.)

Beispiel: Verzeichnungsebene

Data Preparation Tool

Weiterentwicklung: Unterstützung für EAD 3.0

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Direkte Unterstützung des internationalen Standards **EAD 3.0**
  - EAD2002 bereits neben EAD(DDB) unterstützt
  - Herausforderung: sehr flexible Nutzung und hierarchische Verschachtelung von Elementen, offene Granularität im Vergleich zu EAD(DDB)
  - Lösung: Connector mit einem grundlegenden Mapping, welches sukzessive erweitert wird, wenn neue Anforderungen kommen
    - so bereits für Lieferungen in EAD2002 umgesetzt

Data Preparation Tool

Ausblick

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Zeitnahe Veröffentlichung als Open Source
- Bereitstellung insbesondere für …
    - Datengeber zur Aufbereitung und Validierung Ihrer Daten
    - Aggregatoren und Softwarehersteller zur Unterstützung der Schnittstellenentwicklung
- Technische sowie Benutzerdokumentation
- Ermöglicht sukzessive Anpassung an neue Anforderungen

Data Preparation Tool

Fazit

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Die Lieferung an Portale stellt für viele Einrichtungen nach wie vor eine Herausforderung dar, da …
  - … sich die **heterogene Erschließung** nicht immer verlustfrei in den Portalen abbilden lässt
  - … die Datengeber stark von den Möglichkeiten der genutzten **Softwareinfrastruktur** abhängig sind
  - … die Nutzung weitergehender Features (z.B. die Verwendung des DFG-Viewers) z.T. eine Überarbeitung der eigenen Infrastruktur voraussetzt

Data Preparation Tool

Fazit

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

- Die **komplexe Struktur** von EAD, insbesondere der hierarchische Aufbau, bietet viel Potenzial für Fehler und setzt eine **Persistenz** der Datenhaltung voraus, die bei den Einrichtungen nicht immer gegeben ist
- Teilnahme an Portalen soll möglichst vielen Archiven ermöglicht werden
- Data Preparation Tool als nachhaltige Lösung zur Konsolidierung sowie Evaluation/Validierung von Exportdaten

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Vielen Dank
für Ihre Aufmerksamkeit.

… Fragen?

Oliver Götze
Landesarchiv Baden-Württemberg
DDB-Fachstelle Archiv / Archivportal-D
oliver.goetze@la-bw.de / archiv@deutsche-digitale-bibliothek.de

Bildnachweis

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Python-Logo. Python Software Foundation. (Folie 18)

https://www.python.org/community/logos/

lxml-Logo. Stefan Behnel: Implementing XML languages with lxml. (Folie 18)

https://lxml.de/s5/lxml-ep2008.html

requests-Logo. (Folie 18)

http://docs.python-requests.org/en/master/

Qt-Logo. Qt Project. (Folie 18)

https://de.wikipedia.org/wiki/Datei:Qt_logo_2016.svg

Bulma Logo. (Folie 18)

https://github.com/jgthms/bulma