

**DATENQUALITÄT MESSEN
DANK OFFENEN STANDARDS
UND LINKED OPEN DATA**

**DATENQUALITÄT MESSEN
DANK OFFENEN STANDARDS
UND LINKED OPEN DATA**

MUSEUM-DIGITAL:QA

MUSEUM-DIGITAL



MUSEUM-DIGITAL

- Daten von über 1000 Museen gemeinsam veröffentlicht
- Mehrere Hundert verwenden musdb zum Inventarisieren
- Eine gute Qualität der Daten ist entscheidend, und musdb versucht dazu anzuregen
 - Per einfacher (fast unsichtbarer) Verknüpfung mit LOD
 - Und: Tools zur Prüfung von Vollständigkeit und Plausibilität der Angaben zu Objekten



PUQI

- Ermittelt einen numerischen Wert der Datenqualität für die Publikation
 - Anzeige in musdb mittels Rot / Blau / Grün gefärbten Balken
- Checkt v.a. Vorhandensein von Inhalten
 - Aber z.B. auch Lizenzstatus
 - Bildgrößen
 - Länge von Beschreibungstexten



PUQI

- Hier: Gute Erfassung
- Balken: Grün, vollständig ausgefüllt
- Viel Lob (“Gut so”)

Qualitätsindex



Gut so

528 Zeichen in der Objektbeschreibung. Das ist gut!

Dem Objekt wurde mindestens ein Ereignis zugeordnet.

5 Schlagworte/Bezüge eingegeben. (Ideal: 3-9)

Das Objekt wurde einer Objektgruppe (Objektgruppen) zugeordnet

Ein oder mehr Transkripte wurden für das Objekt erstellt

Dem Objekt wurden mehrere Bilder zugeordnet

Bild ist groß (6)

Metadaten: Eine offene Lizenz wurde gewählt



PUQI

- Hier: Sehr rudimentäre Erfassung
- Balken: Rot, nur minimal ausgefüllt
- Viele Verbesserungsvorschläge
(“Könnte evtl. verbessert werden”),
ein wenig Lob (“Gut so”)

Qualitätsindex



Könnte eventuell verbessert werden

Objektname aus einem einzigen häufig vorkommenden Wort (8 Objekte gleichen Namens)

Doppelung: Exakt die gleiche Objektbeschreibung wird auch für andere Objekte verwendet!

8 Zeichen in der Objektbeschreibung. Das ist zu kurz!

Feld "Material/Technik" ist leer. Besser, wenn ein Hinweis dort steht.

Feld "Maße" ist leer. Besser, wenn ein Hinweis dort steht.

Es wurde nur ein Schlagwort/Bezug eingegeben. (Ideal: 3-9)

ADD

Zum Objekt wurde noch kein Bild hochgeladen.

ADD

Gut so

Dem Objekt wurde mindestens ein Ereignis zugeordnet.

Metadaten: Eine offene Lizenz wurde gewählt

PLAUSIBILITÄTSPRÜFUNG

- Vergleicht dem Objekt zugeordnete Ereignisse (Herstellung, Nutzung, etc.)
 - V.a. auch auf Basis von Lebensdaten, bezogen etwa aus Wikidata, GND
- Erkennt unlogisch eingetragene Beziehungen, z.B.:
 - **Heiner Geißler** hat das Objekt **1855** hergestellt
 - **Heiner Geißler** hat das Objekt hergestellt, **Christoph Kolumbus** hat es **besessen**
 - Objekt wurde **1810** hergestellt, **1720** genutzt
 - Objekt wurde **1820** konzeptioniert, **1720** hergestellt



PLAUSIBILITÄTSPRÜFUNG

- Vergleicht dem Objekt zugeordnete Ereignisse (Herstellung, Nutzung, etc.)
 - V.a. auch auf Basis von Lebensdaten, bezogen etwa aus Wikidata, GND
- Erkennt unlogisch eingetragene Beziehungen, z.B.:
 - ***Heiner Geißler hat das Objekt nach 1855 hergestellt***
(Möglicherweise logisch, möglicherweise nicht)



- Erzeuge PDF-Ausgabe
- Verberge DM-Buttons
- Verberge das Objekt
- Objekt exportieren
- Untitled Document 1.htm
- Test report
- Test report
- Test report
- Test report
- Test report
- Test report
- Test report
- Objekt löschen

unterworfen wird. Da die Sehne vor langer Zeit entfernt wurde, konnte sich der Bogen in seine ursprüngliche Form zurückbiegen.

In württembergischen Besitz gelangten die Bögen im Zusammenhang mit den Türkenkriegen. Diese Auseinandersetzungen verschiedener westeuropäischer Staaten mit dem Osmanischen Reich begannen mit der Belagerung Wiens im Jahr 1529 und

Material / Technik: Horn, Leder, Holz

Maße: Höhe: 85,3 cm

ABSCHICKEN

- Hergestellt (1855) vor Geburt von Heiner Geißler
- Wurde genutzt (Akteur: Theoderich der Große (451-526) (* 451)) vor Hergestellt (1855)

- Sammlung

- Kunst- und Kulturgeschichtliche Sammlungen
- Waffen und Militaria

+ Ereignis

Hergestellt Wer: Heiner Geißler Wo: Türkei Wann: 1855	Wurde genutzt Wer: Theoderich der Große Wo: ----- Wann: -----
---	---

- Schlagwort / Bezug

- Bögen (Objektart)
- Horn (Material)

NEU: PLAUSIBILITÄTSPRÜFUNG FÜR BILDLICENSEN

- Prüft auf Basis der Herstellung eines Objektes, ob Bild-Lizenzangaben logisch sind
- Grundannahme: Objektabbildungen wurden zu dokumentarischen Zwecken erstellt
D.h.: Keine eigene Schöpfungshöhe
- Vertragliche Sonderregelungen können immer anderes Vorschreiben, Check ist also nur grobe Annäherung



- Erzeuge PDF-Ausgabe
- Verberge DM-Buttons
- Verberge das Objekt
- Objekt exportieren
- Untitled Document 1.htm
- Test report
- Test report
- Test report
- Test report
- Test report
- Test report
- Objekt löschen

Material / Technik

Horn, Leder, Holz

Maße

Höhe: 85,3 cm

ABSCHICKEN

- **Abbildung "Reflexbogen (#56364)":**
Die letzten eingetragenen Schaffenden des Objektes sind seit 100 oder mehr Jahren tot. Das Objekt und seine (dokumentarischen) Abbildungen sind wahrscheinlich Gemeinfrei (Lizenz-Status: Public Domain Mark). Derzeitiger Lizenz-Status: CC BY-NC-SA.
- **Abbildung "Reflexbogen (#56365)":**
Die letzten eingetragenen Schaffenden des Objektes sind seit 100 oder mehr Jahren tot. Das Objekt und seine (dokumentarischen) Abbildungen sind wahrscheinlich Gemeinfrei (Lizenz-Status: Public Domain Mark). Derzeitiger Lizenz-Status: CC BY-SA.

+ Ereignis

Wurde genutzt
Wer: [≡] Friedrich von Buch

Gedruckt
Wer: [≡] Paracelsus



- Erzeuge PDF-Ausgabe
- Verberge DM-Buttons
- Verberge das Objekt
- Objekt exportieren
- Untitled Document 1.htm
- Test report
- Test report
- Test report
- Test report
- Test report
- Test report
- Test report
- Objekt löschen

Material / Technik

Horn, Leder, Holz

Maße

Höhe: 85,3 cm

ursprüngliche Form zurücklegen

In württembergischen Besitz gelangten die Bögen im Zusammenhang mit den Türkenkriegen. Diese Auseinandersetzungen verschiedener westeuropäischer Stämme mit dem Osmanischen Reich begannen mit der Belagerung Wiens im Jahr 1529 und

- Abbildung "Reflexbogen (#56364)":

Die letzten eingetragenen Schaffenden des Objektes leben noch oder sind in den letzten 50 Jahren gestorben. Das Objekt und seine (dokumentarischen) Abbildungen sind wahrscheinlich noch eingeschränkt ("Alle Rechte vorbehalten"). Derzeitiger Lizenz-Status:: CC BY-NC-SA. Die Herstellenden oder jemand gleichnamiges wird von einer Verwertungsgesellschaft vertreten: VG Bildkunst.

- Abbildung "Reflexbogen (#56365)":

Die letzten eingetragenen Schaffenden des Objektes leben noch oder sind in den letzten 50 Jahren gestorben. Das Objekt und seine (dokumentarischen) Abbildungen sind wahrscheinlich noch eingeschränkt ("Alle Rechte vorbehalten"). Derzeitiger Lizenz-Status:: CC BY-SA. Die Herstellenden oder jemand gleichnamiges wird von einer Verwertungsgesellschaft vertreten: VG Bildkunst.

Wurde genutzt
Wer: [≡] Friedrich von Buch
Wo: ----
Wann: ----

Gedruckt
Wer: [≡] Niki de Saint Phalle
Wo: [≡] Türkei
Wann: [≡] 1855

Hilft das?

Reicht das?

Ja!

Nein!



Hilft das?

Reicht das?

Ja!

Nein!

**Nicht alle Museen
inventarisieren mit
musdb**



ICH HABE EINEN TRAUM

DASS EINES TAGES

**ALLE MUSEEN BESSERE
DATEN HABEN**

ICH HATTE EINE IDEE

**ALLEN MUSEEN,
UNABHÄNGIG VON IHRER
EINGESETZTEN SOFTWARE,
DIE TOOLS AUS MUSDB
ZUR VERFÜGUNG ZU STELLEN**

WIE?



MUSEUM-DIGITAL:QA

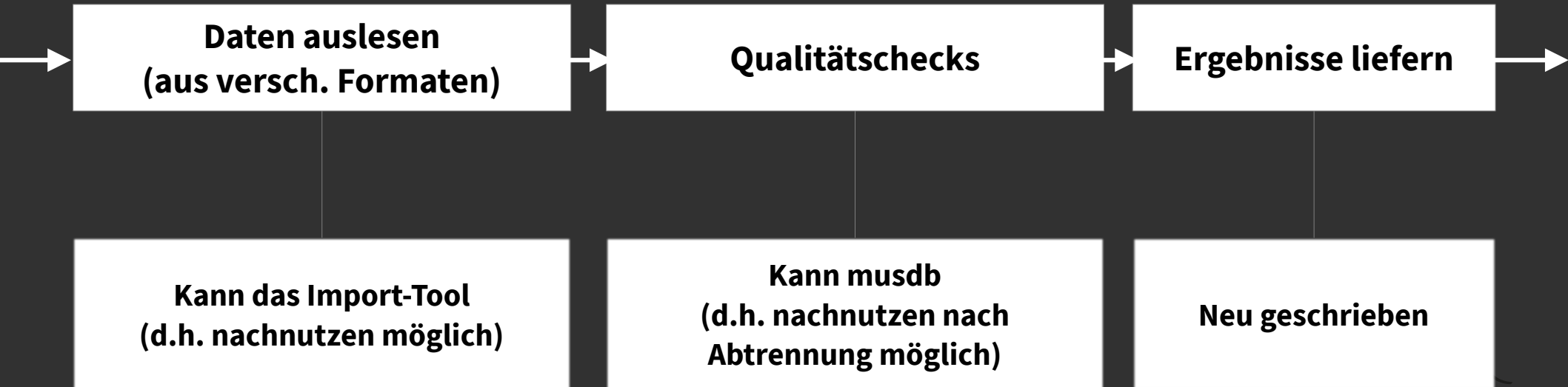
**Daten auslesen
(aus versch. Formaten)**

Qualitätschecks

Ergebnisse liefern



MUSEUM-DIGITAL:QA



MUSEUMSDATEN IMPORTIEREN



DATEN IMPORTIEREN

LIDO

Problem

Es gibt viele verschiedene Eingabeformate und Ursprungs-Programme

musdb

Axiell Collections

Primus

FirstRumos

MuseumPlus



IMPORTIEREN, DER “EINFACHE” WEG

LIDO

musdb

Axiell Collections

Primus

FirstRumos

MuseumPlus

Importieren

- 1) Datenbankverbindung auf
- 2) 1. Objekt
 - 1) Titel auslesen, direkt in DB schreiben
 - 2) Schlagworte auslesen, direkt in DB schreiben
 - 3) (...)
- 3) 2. Objekt
 - 1) (Siehe Oben)
- 4) Datenbankverbindung zu



DER “EINFACHE” WEG (EIN MODUL)

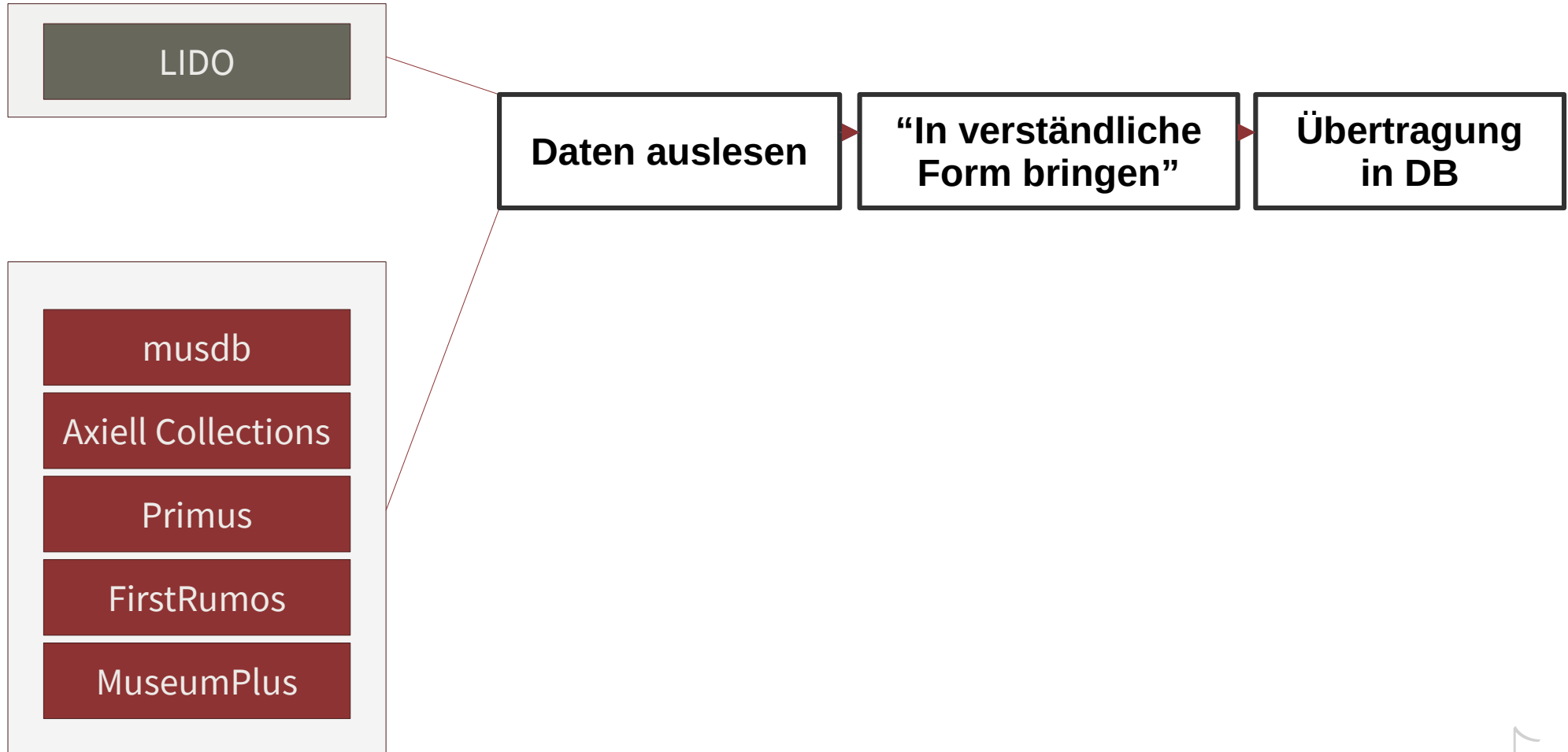
- Ist für Programmierer am schnellsten zu schreiben
- Importprozess ist jeweils spezifisch auf ein Importformat ausgerichtet
- “Spagetti-Code” / “Scripten statt Programmieren”

(Kann manchmal sinnvoll sein – aber nur, wenn man ihn definitiv nicht nochmal braucht)

Kurzfristig verlockend, langfristig teuer



IMPORTIEREN, DER ELEGANTERE WEG



LIDO

Daten auslesen

“In verständliche Form bringen”

Übertragung in DB

musdb

Axiell Collections

Primus

FirstRumos

MuseumPlus

- Ist für jedes Import-Format spezifisch
- Verallgemeinerung max. auf Ebene von Dateiformaten (alle CSV-Dateien sind zeilenbasiert, XML nicht)

- Definition von grundlegenden / allgemeinen Datentypen
(Objekt, Ausstellung, Ereignis)
- Inkl. Funktionen zur einfacheren Ansprache
(“gib mir alle verknüpften Schlagworte des Objektes”)
- Inkl. Abgleich auf Basis von Normdatenbezügen

- Gebündeltes Schreiben in die Datenbank



DER ELEGANTE WEG (DREI MODULE)

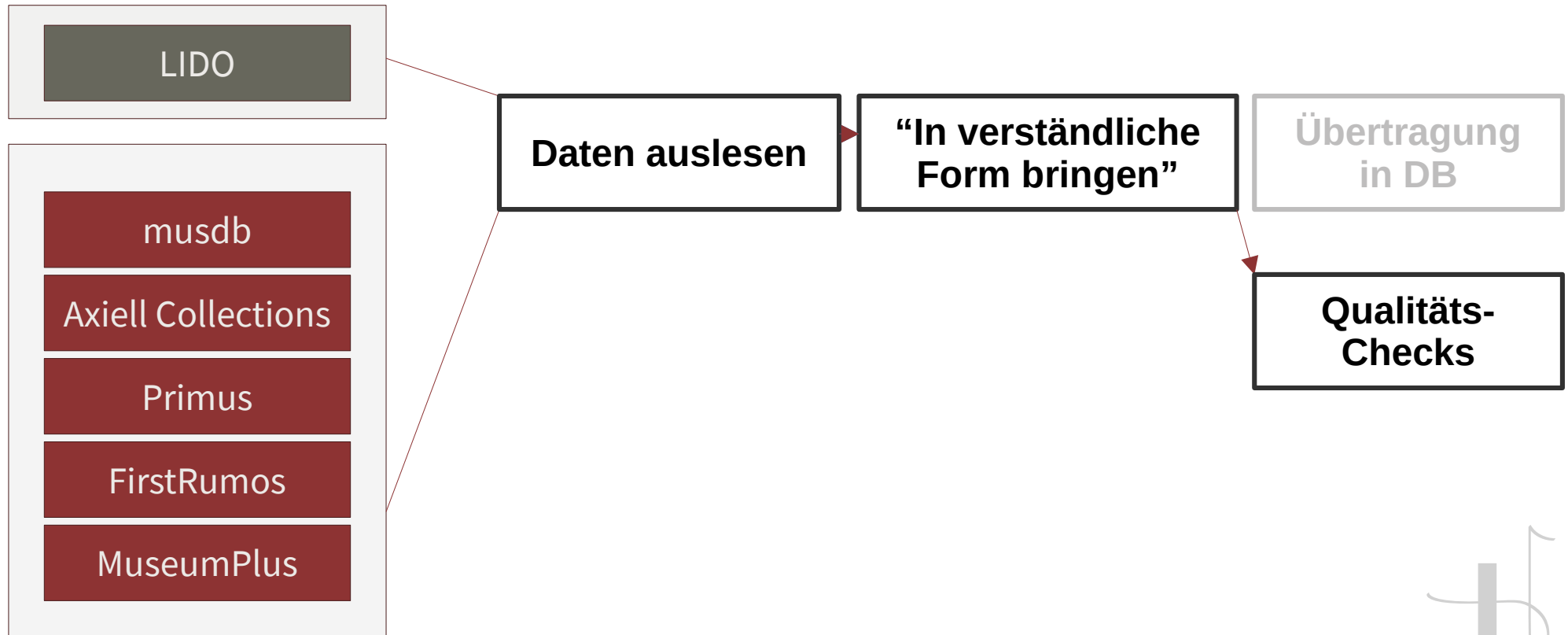
- Nur der Auslese-Vorgang ist formatspezifisch
- Definition der Datentypen und Schreiben muss für alle Arten von Importdaten nur noch einmal implementiert werden

Kurzfristig mehr Arbeit, langfristig deutlich sparsamer

Und bei museum-digital nach einem Rewrite des Import-Tools 2019 ziemlich strikt umgesetzt, jetzt komplett



WEITERER VORTEIL: ANDERE VERWENDUNG STATT SCHREIBEN / IMPORTIEREN



**VERSCHIEDENE
FORMATE AUSLESEN**

**DANN AUF
QUALITÄT PRÜFEN**

MUSEUM-DIGITAL:QA



<https://quality.museum-digital.org/>





Ausprobieren

Ein Import-Format auswählen

> AdlibCsv	>
> AdlibXml	>
> Csvxml	>
> CsvxmlCsv	>
> FirstRumosCsv	>
> JohannHilf	>
> Lido	>
> Primus	>

WERKZEUGE ZUR DATEN-QUALITÄTSKONTROLLE

Seit den Anfängen in den 1980er Jahren inventarisieren mehr und mehr Museen ihre Bestände digital. Wo vorher auf Inventarkarten oft nur die grundlegendsten Informationen zum Objekt systematisch erfasst wurden, erlaubt die digitale Erfassung eine deutlich detailliertere Beschäftigung mit den Beständen ohne die dem Medium der Inventarkarte inhärenten Platzbegrenzungen.

Detailinformationen, die früher in Katalogen oder Fachliteratur veröffentlicht und damit nur abseits der Inventarisierung und in oft unsystematischer Form vorlagen, lassen sich digital gemeinsam mit den grundlegenden Daten der Objekte speichern und durchsuchen. Parallel dazu veröffentlichen Museen ihre Bestände zunehmend in frei durchsuchbaren Datenbanken, oft auch gemeinsam mit anderen Museen die gänzlich andere Sammlungsfoki und Bestände haben.

Durch die Digitalisierung wird die hochqualitative und gleichzeitig systematische Erschließung der Bestände also erst möglich. Durch die oft mit der Erfassung eng verzahnte Veröffentlichung ist die Qualität der Objektdaten andererseits heute relevanter denn je.

museum-digital ermöglicht Museen und verwandten Einrichtungen seit 2009 das gemeinsame Verwalten und Veröffentlichen ihrer Bestände. In diesem Rahmen entstanden Tools zur Messung und Verbesserung der Datenqualität.

Mit museum-digital:qa werden diese Werkzeuge auch über die Grenzen von museum-digital hinaus zugänglich und nachnutzbar gemacht. Dazu können sie einerseits direkt im Web benutzt werden - andererseits über eine Programmierschnittstelle. Über diese können die Qualitätschecks von museum-digital zukünftig auch leicht in andere Sammlungsmanagement-Systeme, die bisher keine vergleichbaren Werkzeuge anbieten, integriert werden.



Ausprobieren

Ein Import-Format auswählen

- > AdlibCsv >
- > AdlibXml >
- > Csvxml >
- > CsvxmlCsv >
- > FirstRumosCsv >
- > JohannHilf >
- > Lido >
- > Primus >

Ausprobieren

EINE DATEI HOCHLADEN

ALS TEXT EINFÜGEN

ZURÜCK

Ausprobieren

Eine Datei hochladen

Datei zum Hochladen auswählen

Browse... No file selected.

ZURÜCK

WERKZEUGE

Seit den Anfängen der Digitalisierung werden Bestände erfasst und in Datenbanken gespeichert. Wo vorher nur die Bestände erfasst wurden, werden heute auch Detailinformationen wie Inventarisierung, Inventarlegende, Grundlegendes über ihre Bestände und die Museen die sie

CONTROLLE

mehr und mehr Museen sammeln Informationen über ihre Bestände. Diese Informationen sind oft detaillierter als die in den Datenbanken. Parallel dazu werden die Datenbanken, oft auch durch die Museen, aktualisiert.

Durch die Digitalisierung wird die hochqualitative und gleichzeitig systematische Erfassung von Beständen also erst möglich. Durch die oft mit der Erfassung eng verzahnte Veröffentlichung ist die Qualität der Objektdaten andererseits heute relevanter denn je.

museum-digital ermöglicht Museen und verwandten Einrichtungen seit 2009 das gemeinsame Verwalten und Veröffentlichen ihrer Bestände. In diesem Rahmen entstanden Tools zur Messung und Verbesserung der Datenqualität.

Mit museum-digital:qa werden diese Werkzeuge auch über die Grenzen von museum-digital hinaus zugänglich und nachnutzbar gemacht. Dazu können sie einerseits direkt im Web benutzt werden - andererseits über eine Programmierschnittstelle. Über diese können die Qualitätschecks von museum-digital zukünftig auch leicht in andere Sammlungsmanagement-Systeme, die bisher keine vergleichbaren Werkzeuge anbieten, integriert werden.

1 Objekte identifiziert

X

Ergebnisse

Inventarnummer: 10-3-3-1

PuQI-Wert	138
Plausibilitäts-Check: Status	Prüfung bestanden
Plausibilitäts-Check für Rechteangaben von Objektabbildungen: Status	Prüfung bestanden

PuQI-Meldungen

- Objektname besteht aus einem einzelnen Wort. Das ist möglicherweise zu unspezifisch
- 171 Zeichen in der Objektbeschreibung. Das ist recht kurz!
- Dem Objekt wurde mindestens ein Ereignis zugeordnet.
- Es wurde nur ein Schlagwort/Bezug eingegeben. (Ideal: 3-9)
- Das Objekt wurde keiner Sammlung zugeordnet.
- Zum Objekt wurde noch kein Bild hochgeladen.
- Beschriftung wurde erfasst und ist öffentlich
- Metadaten: Eine restriktive Lizenz wurde gewählt

FÜR ALLE

1 objects identified

X

Results

Inventory number: 10-3-3-1

PuQI Score	138
Plausibility Check: Status	Check passed
Plausibility Check for License Status of Images / Representations: Status	Check passed

PuQI Messages

- Object title consist of only one word. This might be too unspecific
- 171 Characters used for object description. That is quite short!
- At least one event assigned to the object.
- Only one keyword (or general assignment) provided. (Ideal: 3-9)
- The object was not assigned to a collection.
- No image has been uploaded for the object yet
- Inscription has been recorded and is public.
- Metadata: A restrictive licence is used

Swagger UI

/api

Explore

museum-digital Quality Assessment API 0.0.1 OAS3

API

This documentation describes museum-digital's public API for quality assessments of object data.

[Contact the developer](#)

[Handbook for museum-digital \(German\)](#)

Servers

/api

evaluation

POST /evaluate/object/ Evaluates the quality of an object record

Parameters

Try it out

No parameters

Request body required

application/x-www-form-urlencoded

parser required
string Format by which to parse the input data.

data required
string Object records to parse in the format specified by the parser variable.

lang
string Language, in which the results should be returned.

Responses

Code	Description	Links
200	Returns the ID of the new object and potential errors	No links
	Media type: <input type="text" value="application/json"/> Controls Accept header: Example Value Schema	
	<pre>{ "results": [{ "idnum": "Ms-12345", "puqi": { "score": 0, "msg": [{ "action": "objectName", "status": "warning", "message": "string", "score": 0 }] }, "plausi": { "warn": true, "msg": ["string"] }, "plausi_legal": { "warn": true, "msg": ["string"] } }] }</pre>	
403	Missing permissions	No links
404	Non-existing object or field queried	No links

DAS HEIßT:

WAS BRINGT MD:QA?

**Plattform zum Auslesen von Daten
aus einer Vielzahl von Formaten**

**Qualitätschecks wie in musdb, jetzt
allgemein verfügbar**

**Verfügbar über Weboberfläche oder
per API**

FÜR ALLE

**Mehrsprachigkeit:
Verfügbar bisher in Deutsch und
Englisch**

**Offene REST-API mit OpenAPI-
Definition**



WOHIN GEHT DIE REISE?

Sie alle:

**Integration in Workflows / Daten überprüfen /
md:qa ausprobieren**



WOHIN GEHT DIE REISE?

Sie alle:

Integration in Workflows / Daten überprüfen / md:qa ausprobieren

Software-Anbieter:

**Möglichkeit zur Integration in andere
Sammlungsmanagement-Systeme**



WOHIN GEHT DIE REISE?

Sie alle:

Integration in Workflows / Daten überprüfen / md:qa ausprobieren

Software-Anbieter:

**Möglichkeit zur Integration in andere
Sammlungsmanagement-Systeme**

museum-digital:

**Integration von weiteren Tools in museum-
digital:qa**





Integration von weiteren Tools in museum-digital:qa ...

**Wie wäre es z.B. mit
autogenerierten
Objektbeschreibungen...
mit Llamas und Alpacas**



Wenigstens die Feldliste der AG Minimaldatensatz geht schonmal

Ausprobieren

Bitte wählen Sie die Art der auszuführenden Prüfung

REGULÄR

AG MINIMALDATENSATZ

ZURÜCK

1 Objekte identifiziert

X

Ergebnisse

Inventarnummer: 10-3-3-1

- Das Feld "Objekttitel oder -benennung" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Objekttyp oder -bezeichnung" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Themenkategorie" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Inventarnummer" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Objektbeschreibung" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Material" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Technik" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Maße" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Ereignis in der Objektgeschichte (Feldgruppe)" ist vorhanden und ausgefüllt. Gut.
- Das Feld "Inhaltsschlagwort" ist vorhanden und ausgefüllt. Gut.
- Das Pflichtfeld "Mediendatei (Feldgruppe)" ist nicht vorhanden oder nicht ausgefüllt.
- Das Pflichtfeld "Nutzungsrechte Mediendatei" ist nicht vorhanden oder nicht ausgefüllt.
- Das Pflichtfeld "Rechteinhaber*in Mediendatei" ist nicht vorhanden oder nicht ausgefüllt.

museum-digital Quality Assessment API

0.0.1 OAS3

/api

evaluation

POST /evaluate/object/ Evaluates the quality of an object record

minimaldatensatz

POST /evaluate/minimaldatensatz/ Evaluates the conformity of an object record with the field set of recommended and required fields according to the AG Minimaldatensatz.

Parameters

Try it out

Vielen Dank!



JOSHUA RAMON ENSLIN

jenslin@freies-deutsches-hochstift.de

Herbsttagung 2023 der Fachgruppe Dokumentation des Deutschen Museumsbundes

9.10.2023 - 11.10.2023

Konrad-Zuse-Zentrum Berlin, Takustraße 7

Vortrag am 10.10.2023, 15:45-16:15

