

Eine Online-Abfrageschnittstelle für die Expo-DB

(13.12.2018, Version 4)

Die Expo-DB enthält XML-Datensätze, die aus imdas-pro-Datenbanken regelmäßig automatisiert exportiert werden. Für die Abfrage der Daten aus der Expo-DB sind diese Datensätze mit einer Reihe von Attributen und Indexen versehen, die eine Identifizierung sowie Selektieren und Sortieren ermöglichen.

1. Der Aufbau der URL zur Datenabfrage

Die Online-Schnittstelle zur Expo-DB wird mit dem Protokoll https angesprochen. Für die folgenden Beispiele wird die (fiktive) Domäne expo.bsz-bw.de verwendet; diese wird je nach Installation der Schnittstelle anders lauten. Die Basis-Urls zu ihren Expo-Db-Schnittstellen würde entsprechend <https://expo.bsz-bw.de/r0/> lauten (die Pfadkomponente „r0“ steht für Release 0 oder Beta-Release; künftige Fortentwicklungen der Schnittstellen werden entsprechend unter „r1“, „r2“,... erreichbar sein). Die Schnittstelle ist mandantenfähig wobei die Mandanten derzeit über eine Nutzerkennung ausgewählt werden; dies könnte zukünftig durch einen URL-Komponente abgeöst werden.

2. Der Abruf von Einzeltreffern

Ein Einzeltreffer wird durch das URL-Suffix /detail anhand eines obligatorischen id-Parameters abgerufen und kann mit Hilfe eines optional im fmt-Parameter angegebenen Formats transformiert werden. Sofern kein fmt-Parameter angegeben wird, wird der Einzeltreffer in einem XML-Format ausgegeben, das wesentliche Datenelemente enthält.

Bezeichnung	Beschreibung
id	Dieser Parameter bezeichnet die Id eines Einzeltreffers, wobei der Wert die imdas-id bezeichnet.
fmt	Dieser Parameter bezeichnet das Format, in dem die Treffer zurückgegeben werden. Die möglichen Formate und ihre Bezeichnungen werden pro Schnittstelle vorab festgelegt (Vorgabewert: XML-Format, das wesentliche Datenelemente enthält).

Die Rückgabe des Einzeltreffers erfolgt in einem https-Response, das je nach im fmt-Parameter gewählten Format den Mimetype application/xml oder application/json aufweist. Der Datensatz wird in einem XML- bzw. JSON-Element „record“ eingeschlossen.

Beispiel:

<https://expo.bsz-bw.de/r1/detail?id=000FDDCB46B9491EDA22F1BDDCB46B5B&fmt=jsondefault>

```
{ „record“ : {
  „sammlung“ : „Moderne Malerei“,
  „kuenstler“ : „Max Müller“,
  „titel“ : „Ein Bild“,
  „material“ : „Leinwand“,
  „technik“ : „Ölmalerei“,
```

```

    „entstehungsjahr“ : „2017“
  }
}

```

3. Der Abruf von Trefferlisten

Für den Abruf von Trefferlisten erfolgt über das Suffix /liste und es sind folgende Parameter vorgesehen, die alle optional sind und, sofern sie fehlen, mit Vorgabewerten belegt werden. Die Reihenfolge, in der die Parameter angegeben werden, ist nicht signifikant. Sie werden im Querystring der URL als name=value-Paare, die mit & getrennt sind, codiert. Zur Darstellung von reservierten Zeichen (z.B. Leerzeichen) und Sonderzeichen ist ein URL-Encoding der Parameterwerte erforderlich.

Bezeichnung	Beschreibung
coll	Dieser Parameter bezeichnet eine Teilsammlung, auf die eine Treffermenge beschränkt werden soll. Die Bezeichnungen für Teilsammlungen müssen vorab festgelegt werden und können einer hierarchischen Struktur folgen, so dass z.B. alle Datensätze in der Teilsammlung mit der Bezeichnung „mus.malerei.moderne“ auch in der Teilsammlung mit der Bezeichnung „mus.malerei“ enthalten sind.(Vorgabewert: keine Beschränkung).
qry	Dieser Parameter bezeichnet eine Selektion der Treffermenge. Die dafür verwendete Anfrage-Syntax wird unten beschrieben (Vorgabewert: alles).
srt	Bezeichnet die Sortierordnung, nach der die Treffermenge sortiert wird. Die möglichen Sortierordnungen werden pro Schnittstelle festgelegt. (Vorgabewert: inventarnummer)
fst	Dieser Parameter bezeichnet den ersten Datensatz in der sortierten Treffermenge, der zurückgegeben wird (Vorgabewert: 1).
len	Dieser Parameter bezeichnet die Anzahl der Treffer, die soweit vorhanden, zurückgegeben werden (Vorgabewert: 12)
fmt	Dieser Parameter bezeichnet das Format, in dem die Treffer zurückgegeben werden. Die möglichen Formate und ihre Bezeichnungen werden pro Schnittstelle vorab festgelegt. (Vorgabewert: Ein XML-Format, in dem lediglich eine Datensatz-Id und eine Kurzdarstellung, z.B. Künstler und Titel, enthalten sind.).

Die Rückgabe der Daten erfolgt in einem https-Response, das je nach im fmt-Parameter gewählten Format den Mimetype application/xml oder application/json aufweist.

Zur Rückgabe wird die entsprechend fmt-formatierten Treffer in ein XML- oder JSON-Element „records“ eingeschlossen, dem ein „head“-Element vorangestellt wird. Im Header werden die Trefferanzahl, die Selektion und Sortierung, die Nummer des ersten Treffers sowie das gewählte Transformationsformat angegeben. Gibt es keine Treffer, wird im Header die Trefferanzahl 0 angegeben und das „records“-Element entfällt. Das gesamte Resultat wird in ein „result“-Element eingebettet.

Beispiel:

[https://expo.bsz-bw.de/r1/liste?qry=person+any+\"müller\"](https://expo.bsz-bw.de/r1/liste?qry=person+any+\)

```

<?xml version='1.0' encoding='utf-8'?>
<result>
  <head>
    <cll>mus.malerei.moderne</cll>

```

```

    <qry>person any müller</qry>
    <srt>inventarnummer</srt>
    <fst>1</fst>
    <len>12</fst>
    <fmt>default</fmt>
    <reccount>3</reccount>
</head>
<records>
  <record>
    <id>000FDDCB46B9491EDA22F1BDDCB46B5B</id>
    <title>Max Müller: Ein Bild<title>
  </record>
  <record>
    <id>000FDDCB46B9491EDA22F1BBFD2F435B</id>
    <title>Max Müller: Noch ein Bild<title>
  </record>
  <record>
    <id>000FDDCB46B9491EAFDDCB4BFD2F435B</id>
    <title>Sabine Müller: Auch ein Bild<title>
  </record>
</records>
</result>

```

4. Eine Anfragesprache für die Expo-DB

Zur Formulierung des Selektionskriteriums im qry-Parameter wird eine Syntax festgelegt. Die abzufragende Datenbasis wird als Text oder Zahlen betrachtet und basiert auf den Indexen der Expo-DB. Suchterme können mit * als Wildcard links- und rechtstrunkiert und als Phrase formuliert werden, Listen von Suchtermen können konjunktiv oder disjunktiv zu Indexen relationiert sein, einzelne Suchterme können auch mit kleinergleich, größergleich oder gleich relationiert werden. Suchklauseln aus Index und Suchtermen können über boole'schen Operatoren verknüpft werden, wobei die Auswertung von links nach rechts erfolgt.

Indexe

Die Indexe sind vorab als eindeutige kurze Bezeichnungen, z.B. **nummer**, **person**, **material**, vereinbart; die Bezeichnung **text** ist für den Index über alle suchbaren Textinhalte reserviert.

Suchterme

Suchterme sind einzelne Worte, **Bronze**, **Eisen**, die durch Leerzeichen getrennt zu Listen zusammengefasst werden, die durch Anführungszeichen begrenzt sind: "**Bronze Eisen**". Dem Wort kann ein * vor- oder nachgestellt werden, was eine Rechtstrunkierung bedeutet, z.B. **Eisen*** sucht nach Eisen, Eisennagel und Eisenbahn. Groß- und Kleinschreibung ist bei Suchtermen nicht signifikant, d.h. Eisen und eisen sind semantisch identisch.

Relationen und Suchklauseln

Indexbezeichnern und Suchtermen bzw. Suchtermlisten werden mit den Relationen **any**, **all** und **phr** zu Suchklauseln gruppiert. Die Semantik von **any** ist dabei, dass der Suchterm bzw. wenigstens ein

Term aus einer Suchtermliste im Index enthalten ist: Die Suchklausel **material any "Eisen Bronze"** selektiert also alle Datensätze, für die im Material-Index entweder Eisen oder Bronze oder Eisen und Bronze enthalten ist; die Suchklausel **material any "Gold"** erfordert, dass im Material-Index Gold verzeichnet ist.

Die Semantik von **all** ist, dass der Suchterm bzw. alle Terme aus einer Suchtermliste im Index enthalten ist: Die Suchklausel **material all Eisen Bronze** selektiert also die Datensätze, für die im Material-Index sowohl Eisen als auch Bronze enthalten sind.

Die Semantik von **phr** ist, dass der Suchterm bzw. die Suchtermliste in ihrer Reihenfolge als Phrase im Index enthalten ist: Die Suchklausel **person phr "Hans Meier"** selektiert also die Datensätze, für die im Personen-Index die Wörter Hans Meier unmittelbar aufeinanderfolgend enthalten ist; die Suchklausel **material phr "Gold"** erfordert, dass im Material-Index Gold verzeichnet ist.

Gibt es nur einen Suchterm sind die Relationen any, all und phr semantisch identisch. Für die Relationen eq, le, ge wird immer nur der erste Suchterm in einer Liste berücksichtigt; typischerweise sollte dieser eine (Jahres-)zahl sein.

Die Semantik von **eq, le** bzw. **ge** ist dann, dass der Suchterm gleich, kleiner bzw. größer als ein Term im Index ist: Die Suchklausel **jahr le 1984** selektiert also alle Datensätze zu Jahren kleiner als oder gleich **1984**, die Suchklausel **jahr eq 1985** Datensätze, die als (Herstellungs)jahr **1985** aufweisen.

Boolsche Verknüpfungen

Suchklauseln können über bool'sche Operatoren **and, or** und **not** zu größeren Abfragen kombiniert werden. Der Operator **and** bedeutet dabei, dass sowohl die Suchklauseln links als auch die Suchklausel rechts vom Operator erfüllt sein müssen: **material any "Eisen" and nummer all "4 7b"** selektiert also die Datensätze, die im Material-Index Eisen aufweisen und im Nummern-Index die Einträge 4 sowie 7b.

Der Operator **or** bedeutet, dass die linke Suchklausel erfüllt ist, oder die rechte Suchklausel oder beide. Enthält der Nummern-Index also den Eintrag 7b, wäre demnach der Ausdruck **material any "Eisen" or nummer any "4 7b"** erfüllt.

Der Operator **not** bedeutet, dass die linke Suchklausel erfüllt sein muss, die rechte aber nicht erfüllt sein darf – logisch also ein „and not“: **material any "Eisen" not nummer all "4 7b"** bedeutet demnach, dass der Material-Index Eisen enthalten muss, der Nummern-Index aber weder 4 noch 7b enthalten darf.

Werden mit and, or oder not mehr als zwei Suchklauseln kombiniert, so werden zunächst die beiden linken Suchklauseln ausgewertet und kombiniert, dann das Ergebnis mit der dritten Klausel u.s.w.: **material all "gold" and "nummer" any 9 or jahr le 1984 not nummer all "4 7b"** selektiert also die Datensätze, für die der Material-Index Gold enthält sowie der Nummern-Index 9, oder aber deren Jahr-Index Werte kleiner oder gleich 1984 enthält; jedenfalls darf aber der Nummern-Index weder 4 noch 7b enthalten.

Klammern zur Variation der Auswertungsreihenfolge bei drei und mehr Suchklauseln sind nicht vorgesehen.

Weitere Beispiele

material any "Eisen Bronze" not material all "Gold Messing" or nummer any 4

Material ist Eisen oder Bronze, aber Material ist weder Gold noch Messing oder die Nummer ist 4.

text any "Eisen Meier" and material all "Messing*" not nummer any "4 7"

Irgendwo im Datensatz kommt Eisen und / oder Meier vor und ein Material beginnt mit Messing, aber die Nummer ist weder 4 noch 7.

material all Holz Eisen not nummer any 4 or person all "Meier Müller-Schmidt"

Der Materialien-Index enthält sowohl Holz als auch Eisen, jedoch der Nummern-Index nicht die 4 oder aber, unabhängig vom vorigen, enthält der Personen-Index sowohl Meier als auch Müller-Schmidt.

5. Abruf von Indexlisten !noch nicht implementiert!

In der ExpoDB sind den Datensätzen zusätzlich zu den Indexen für Suche und Sortierung noch spezielle Indexe für Indexlisten zugeordnet, die den Abruf solcher Listen erlauben, z.B. zur Anzeige von Facetten sowie zur Unterstützung der Eingabe von Suchtermen. Folgende Parameter sind dabei vorgesehen, wobei der Parameter idx obligatorisch ist, alle anderen optional:

Bezeichnung	Beschreibung
coll	Dieser Parameter bezeichnet eine Teilsammlung, auf die eine Indexliste beschränkt werden soll. Die Bezeichnungen für Teilsammlungen müssen vorab festgelegt werden und können einer hierarchischen Struktur folgen, so dass z.B. alle Datensätze in der Teilsammlung mit der Bezeichnung „mus.malerei.moderne“ auch in der Teilsammlung mit der Bezeichnung „mus.malerei“ enthalten sind.(Vorgabewert: keine Beschränkung).
qry	Dieser Parameter bezeichnet eine Selektion der Treffer, auf die die Indexliste eingeschränkt wird. Die dafür verwendete Anfrage-Syntax wird oben beschrieben (Vorgabewert: alles).
idx	dieser Parameter bezeichnet einen vorab festgelegten Index, für den die Indexterme ermittelt werden sollen. Sofern statt der Gesamtindexliste nur Indexterme abgerufen werden sollen, die ein bestimmtes Präfix besitzen, muss dieses Präfix mit Colon getrennt der Indexbezeichnung angehängt werden z.B. idx=material:kupf
srt	Angabe der Sortierung der Indexterme. Mögliche Werte sind „a“ für absteigende lexikalische Sortierung sowie „h“ für Häufigkeit (Vorgabe).
fmt	Das Format der Ausgabe. Mögliche Werte sind xml und json (Vorgabe)

Die Rückgabe der Indexlisten erfolgt in einem https-Response, das je nach im fmt-Parameter gewählten Format den Mimetype application/xml oder application/json aufweist. Der Datensatz wird in einem XML- bzw. JSON-Element „index“ eingeschlossen, dem ein „header“-Element vorangestellt wird. Im Header werden idx-Parameter sowie ggf. die anderen Parameter angegeben. Header- und Index-Element werden in ein result-Element eingeschlossen.

Beispiel:

<https://expo.bsz-bw.de/r1/index?idx=material:k&srt=h&fmt=json>

```
{ „result“: {  
  „header“ : {
```

```

    „idx“ : „material:h“,
    „srt“ : „h“,
    „fmt“ : „json“
  },
  „index“ : {
    {
      „term“ : „keramik“,
      „anzahl“ : 45
    },
    {
      „term“ : „karton“,
      „anzahl“ : 33
    },
    {
      „term“ : „kunststoff“,
      „anzahl“ : 18
    },
    {
      „term“ : „kork“,
      „anzahl“ : 2
    }
  }
}

```

6. Abruf von Medien für die die Expo-DB

In der Expo-DB sind insbesondere auch Pfade zu Bildern und anderen Mediendateien abgelegt. Diese Pfade sind intern, werden in der Regel nicht in den Trefferformaten ausgegeben und können nicht direkt zum Abruf dieser Ressourcen über das Web verwendet werden. Vielmehr erfolgt der Abruf von Bildern sowie anderen Mediendateien über einen https-GET-Request, in dem die id des Datensatzes, optional die Nummer des Bildes in der Liste der zum Datensatz publizierten Bilder sowie optional die Bilddimension als Parameter angegeben sind. Folgende Parameter sind dabei vorgesehen:

Bezeichnung	Beschreibung
id	Die Imdas-Id aus der Imdas Pro-Datenbank (s.o.).
img	Die Position des gewünschten Bildes in der Liste der zum Datensatz publizierten Bilder. Sind z.B. zu einem Objekt fünf Bilder publiziert, bezeichnet img=3 das dritte Bild in dieser Liste (Vorgabewert: „1“).
dim	Die Bilder werden beim Abruf auf die gewünschte Größe in Pixeln skaliert; wie dieser Wert interpretiert wird bestimmt der Parameter mode (Vorgabewert: „100“).
mode	Für den Parameter mode sind die Werte „x“: keine Skalierung, Originalgröße, „h“ dim ist die gewünschte Höhe, „w“: dim ist die gewünschte Breite, sowie „d“: dim ist das größere von Länge und Breite, also ein umfassendes Rechteck (Vorgabewert: „d“)

Beispiel:

<https://expo.bsz-bw.de/image?id=000FDDCB46B9491EDA22F1BDDCB46B5B&dim=200&mode=h>

Bilder werden im JPG-Format unter dem Mime-type image/jpeg ausgeliefert.

7. Fehlermeldungen

Falls eine Parametrisierung fehlerhaft ist, also falsch strukturiert ist, nicht vereinbarte Index-Bezeichner verwendet, ein qry-Parameter entspricht nicht der Syntax der Anfragesprache, beim Abruf von Medien wird eine id angegeben, zu dem kein Datensatz existiert, oder ein img-Parameter, zu dem es kein Bild mit der angegebenen Nummer gibt, wird der Fehlercode http-400 zurückgeliefert.

Zusätzlich wird in einer Html-Seite eine Fehlermeldung ausgeliefert, die das Problem beschreibt.

8. Authentifizierung und Autorisierung

Die Schnittstellen zur Expo-DB sind im Zugriff beschränkt werden. Dazu werden Zugangsdaten mit Username / Passwort eingesetzt. Die Nutzerkennung dient derzeit zur Auswahl des Mandanten. Ein Einzeltrefferaufruf, der die Zugangsdaten mueller / geheim beinhaltet, würde lauten:

<https://mueller:geheim@expo.bsz-bw.de/detail?id=000FDDCB46B9491EDA22F1BDDCB46B5B>

Sicherer wäre allerdings hier eine Authentifizierung per Signifikat zwischen dem Host der Schnittstelle und dem anfragenden Server. Dazu wird zunächst die Mandantenauswahl auf eine URL-Komponente umgestellt.

Werden keine Zugangsdaten übermittelt oder misslingt die Authentifizierung so wird der Fehlercode http-403 zurückgeliefert.

Bemerkung: Diese Spezifikation ist noch nicht implementiert. Insofern kann diese Beschreibung noch Fehler aufweisen und wird noch weiterentwickelt. Hinweise zu Fehlern sind willkommen und werden in künftigen Versionen des Dokuments korrigiert. Erweiterungen der Funktionalität werden dagegen in nachfolgenden Releases der Schnittstelle realisiert, wobei parallel die Vorversionen in Betrieb bleiben um die Funktionsfähigkeit von Clients sicherzustellen.