

# Open Data Portale und deren Inhalte

Ist eine automatisierte Abfrage möglich?

# Aufgabe

- Erstellung einer Übersicht über die Open Data Portale aus Deutschland mindestens
- Ermittlung der darin angebotenen Ressourcen / Daten
- Erstellung einer Übersicht, die interaktiv verwendet werden kann
- Bereitstellung der verwendeten Daten für andere Interessierte

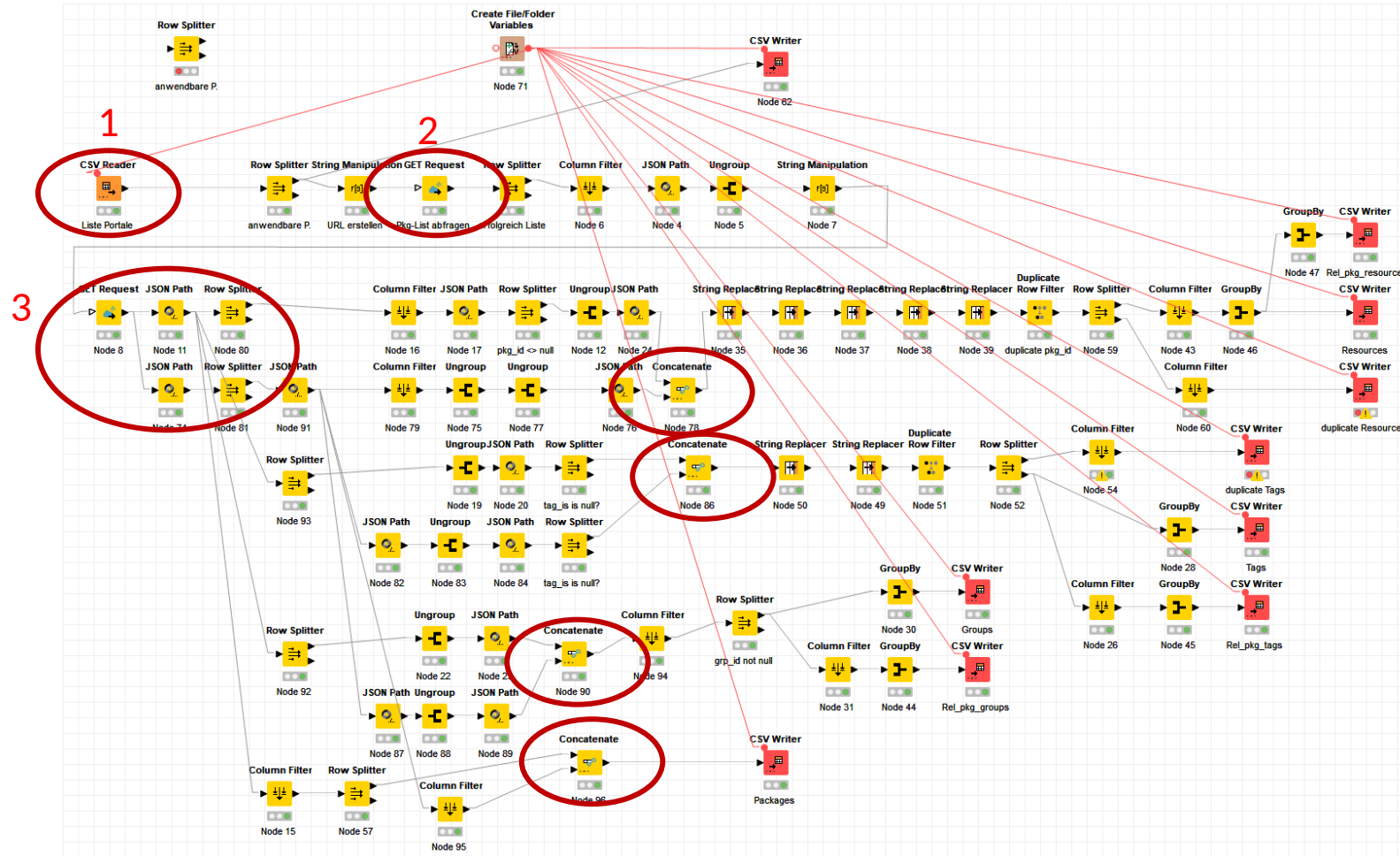
# Schritte

- Erstellung einer Liste der Portale mit ihren URLs und dem Typ des Portals bzw. seiner API zum automatisierten Zugriff.
- Abfrage der Inhalte für CKAN bzw. DKAN Portale (wegen der beschriebenen einheitlichen Schnittstelle und der weiten Verbreitung)
- Ordnen der Inhalte nach den Objekten des Standards
- Abfrage wichtiger Attribute der Objekte
- Abfrage der Beziehungen zwischen Objekten (z.B. welche Tags sind zugeordnet)
- Aufbereitung der Objekte als CSV-Dateien zur weiteren Verarbeitung

# Auswertung

- Erstellung einer Power BI Auswertung auf der Basis der Daten aus den Portalen
- Erstellung eines Datenmodells mit den Objekten und ihren Beziehungen
- Erstellung von Beispielauswertungen

# Datenabruf und Aufbereitung: Übersicht



# Datenabruf und Aufbereitung: Knackpunkte

- In der vorhergehenden Folie wird die Liste der Portale an Punkt 1 eingelesen.
- Die Abfrage an Punkt 2 ermittelt die vom Portal verfügbar gemachten Packages
- Die Abfrage an Punkt 3 ermittelt zu jedem Package aus Punkt 2 die darin enthaltenen Ressourcen
- Hier unterscheiden sich die Ergebnisse der gleichen API-Aufrufe zwischen CKAN und DKAN und es muss eine Fallunterscheidung implementiert werden.
- Die weiteren markierten Punkte führen die Zweige der Fallunterscheidungen wieder zusammen.
- Die roten Komponenten am jeweils rechten Ende der Zweige speichern die aufbereiteten Daten in den Ergebnisdateien.

# Auswertung: Anzahl Artefakte pro Portal

Die Liste links zeigt die Ergebnisse der Auswertung.

Pro Stadt wird sichtbar, wie viele Packages und Resources das zugehörige Portal bereitstellt.

Die Liste rechts zeigt alle Portale in der eingeleseenen Liste. Es wird deutlich, dass nicht bei jedem Portal der Abruf der Packages erfolgreich war.

city	# Pkg	# Resources
Aachen	130	490
Berlin	2.664	8.481
Diverse	34.545	52.748
Duisburg	58	58
Düsseldorf	309	309
Essen	42	42
Frankfurt	99	693
Gelsenkirchen	67	67
Govdata	42.449	116.389
Hamburg	151.315	248.977
HRO	230	2.306
Jena	70	117
Karlsruhe	101	578
Köln	303	303
Konstanz	49	49
Leipzig	710	1.876
Lklg	167	172
Meerbusch	47	174
Mühlheim-Ruhr	36	36
München	169	254
Münster	124	124
OEPNV	28	200
Offenesdatenportal	1.580	2.051
Rlp	4.872	22.308
Ruhr	513	2.505
Schleswig-Holstein	14.613	19.599
Wuppertal	97	97
<b>Total</b>	<b>262.354</b>	<b>481.003</b>

country	city	api
Deutschland	Aachen	<a href="http://offenedaten.aachen.de/api/3/action/package_list">http://offenedaten.aachen.de/api/3/action/package_list</a>
Deutschland	Bayern	<a href="https://opendata.bayern.de/api/1/search/dataset?all_fields=1&amp;q=**">https://opendata.bayern.de/api/1/search/dataset?all_fields=1&amp;q=**</a>
Deutschland	Berlin	<a href="https://datenregister.berlin.de/api/3/action/package_list">https://datenregister.berlin.de/api/3/action/package_list</a>
Deutschland	Bielefeld	<a href="https://open-data.bielefeld.de/api/3/action/package_list">https://open-data.bielefeld.de/api/3/action/package_list</a>
Deutschland	Bonn	<a href="https://opendata.bonn.de/api/3/action/package_list">https://opendata.bonn.de/api/3/action/package_list</a>
Deutschland	Deutschebahn	<a href="https://data.deutschebahn.com/api/3/action/package_list">https://data.deutschebahn.com/api/3/action/package_list</a>
Österreich	Diverse	<a href="http://www.data.gv.at/katalog">www.data.gv.at/katalog</a>
Deutschland	Dormagen	<a href="https://opendata.dormagen.de/api/v2/catalog/datasets">https://opendata.dormagen.de/api/v2/catalog/datasets</a>
Deutschland	Duisburg	<a href="https://opendata-duisburg.de/api/3/action/package_list">https://opendata-duisburg.de/api/3/action/package_list</a>
Deutschland	Düsseldorf	<a href="https://opendata.duesseldorf.de/api/3/action/package_list">https://opendata.duesseldorf.de/api/3/action/package_list</a>
Deutschland	Essen	<a href="https://opendata.essen.de/api/3/action/package_list">https://opendata.essen.de/api/3/action/package_list</a>
Deutschland	Frankfurt	<a href="http://offenedaten.frankfurt.de/api/3/action/package_list">http://offenedaten.frankfurt.de/api/3/action/package_list</a>
Deutschland	Freiburg	<a href="https://fritz.freiburg.de/duva2ckan/files/">https://fritz.freiburg.de/duva2ckan/files/</a>
Deutschland	Gelsenkirchen	<a href="https://opendata.gelsenkirchen.de/api/3/action/package_list">https://opendata.gelsenkirchen.de/api/3/action/package_list</a>
Deutschland	Govdata	<a href="https://ckan.govdata.de/api/3/action/package_list">https://ckan.govdata.de/api/3/action/package_list</a>
Deutschland	Hamburg	<a href="http://suche.transparenz.hamburg.de/api/3/action/package_list">http://suche.transparenz.hamburg.de/api/3/action/package_list</a>
Deutschland	Herne	<a href="https://opendata.herne.de/katalogopendata/doc/api/index.html">https://opendata.herne.de/katalogopendata/doc/api/index.html</a>
Deutschland	HRO	<a href="https://www.opendata-hro.de/api/3/action/package_list">https://www.opendata-hro.de/api/3/action/package_list</a>
Deutschland	Jena	<a href="https://opendata.jena.de/api/3/action/package_list">https://opendata.jena.de/api/3/action/package_list</a>
Deutschland	Karlsruhe	<a href="https://transparenz.karlsruhe.de/api/3/action/package_list">https://transparenz.karlsruhe.de/api/3/action/package_list</a>
Deutschland	Kleve	<a href="https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=stadt-kleve">https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=stadt-kleve</a>
Deutschland	Köln	<a href="https://offenedaten-koeln.de/api/3/action/package_list">https://offenedaten-koeln.de/api/3/action/package_list</a>
Deutschland	Konstanz	<a href="https://offenedaten-konstanz.de/api/3/action/package_list">https://offenedaten-konstanz.de/api/3/action/package_list</a>
Deutschland	Krefeld	<a href="https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=krefeld">https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=krefeld</a>
Deutschland	Kreis-Kleve	<a href="https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=kreis-kleve">https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=kreis-kleve</a>
Deutschland	Kreis-Viersen	<a href="https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=kreis-viersen">https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=kreis-viersen</a>
Deutschland	Leipzig	<a href="https://opendata.leipzig.de/api/3/action/package_list">https://opendata.leipzig.de/api/3/action/package_list</a>
Deutschland	Lklg	<a href="https://open.lklg.net/api/3/action/package_list">https://open.lklg.net/api/3/action/package_list</a>
Deutschland	Mannheim	<a href="https://mannheim.opendatasoft.com/api/v2/catalog/datasets">https://mannheim.opendatasoft.com/api/v2/catalog/datasets</a>
Deutschland	Meerbusch	<a href="https://opendata.meerbusch.de/api/3/action/package_list">https://opendata.meerbusch.de/api/3/action/package_list</a>
Deutschland	Moers	<a href="https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=moers">https://www.offenesdatenportal.de/api/3/action/organization_activity_list?id=moers</a>
Deutschland	Mühlheim-Ruhr	<a href="https://geo.muelheim-ruhr.de/api/3/action/package_list">https://geo.muelheim-ruhr.de/api/3/action/package_list</a>

# Auswertung: Formatangaben der Ressourcen

Die Liste links zeigt, wie viele Ressourcen mit welchem Format aus den API-Aufrufen geliefert werden.

Es fällt auf, dass:

1. Es keine einheitliche Art gibt, das Format anzugeben.
2. Es mehrere Varianten gibt, um z.B. PDF-Dateien zu definieren.
3. Mehrere Angaben bei einer Resource verwendet werden, was nicht nachvollziehbar ist
4. Exotische Ressourcen offenbar auch verfügbar gemacht werden (Webanwendung, Diverse, leeres Format, ...)

# Resources	resource_format	# Resources	resource_resource_type
142.377	PDF	477.648	
101.821	HTML	566	api
57.848	CSV	71	documentation
20.889	GML	1	example
18.426	<a href="http://publications.europa.eu/resource/authority/file-type/CSV">http://publications.europa.eu/resource/authority/file-type/CSV</a>	2.394	file
13.479	WMS	237	file.upload
13.377	ZIP	82	metadata
10.049	Karte	4	visualization
10.048	Webanwendung		
7.046	XML		
6.513	XLSX		
5.459	<a href="http://publications.europa.eu/resource/authority/file-type/PDF">http://publications.europa.eu/resource/authority/file-type/PDF</a>		
4.162	Diverse		
3.798	<a href="http://publications.europa.eu/resource/authority/file-type/XML">http://publications.europa.eu/resource/authority/file-type/XML</a>		
3.587	XLS		
3.515	WFS		
3.131	<a href="http://publications.europa.eu/resource/authority/file-type/JSON">http://publications.europa.eu/resource/authority/file-type/JSON</a>		
2.711			
2.710	download		
2.611	<a href="http://publications.europa.eu/resource/authority/file-type/HTML">http://publications.europa.eu/resource/authority/file-type/HTML</a>		
2.475	tif, png, pdf, gml, dxf, shp		
2.451	<a href="https://www.iana.org/assignments/media-types/text/csv">https://www.iana.org/assignments/media-types/text/csv</a>		
2.444	<a href="http://publications.europa.eu/resource/authority/file-type/XLSX">http://publications.europa.eu/resource/authority/file-type/XLSX</a>		
2.303	view		
2.239	JSON		
2.125	XPlanGML		
1.933	<a href="https://www.iana.org/assignments/media-types/text/xml">https://www.iana.org/assignments/media-types/text/xml</a>		
1.736	<a href="http://publications.europa.eu/resource/authority/file-type/GML">http://publications.europa.eu/resource/authority/file-type/GML</a>		
1.722	SHP		
1.597	<a href="http://publications.europa.eu/resource/authority/file-type/XLS">http://publications.europa.eu/resource/authority/file-type/XLS</a>		
1.575	jpg		
481.003		481.003	



# Auswertung: Formatangaben der Ressourcen

Die Liste zeigt, wie viele Ressourcen aus Hamburg mit dem Tag „Umwelt“ geliefert werden.. Es fällt auf, dass:

1. Es werden Dokumente bereitgestellt, die man nicht bei offenen Daten erwartet hätte.
2. An manchen Dokumenten hätte man das Tag „Umwelt“ nicht erwartet (z.B. „Börsentag Hamburg“)
3. In der Auswertung offenbar recht sprechende Namen für die Tags verwendet werden.

# Pkg	# Resources	pkg_title	tag_name
1	1	- Aussenwanddämmung der Hoffassaden - Abbruch von Kragbalkonen - Errichtung von zwei Vorstellbalkonen	<input type="checkbox"/> Select all <input type="checkbox"/> Energie- und Umwelttechnik <input type="checkbox"/> EnergieundUmwelt <input type="checkbox"/> Informationssystem Stadt und Umwelt <input type="checkbox"/> Investitionen für Umweltschutz im Produzierenden Ge... <input type="checkbox"/> Natur und Umwelt <input type="checkbox"/> schädliche Umwelteinwirkungen <input type="checkbox"/> Strategische Umweltprüfung <input type="checkbox"/> Umweltschutz <input checked="" type="checkbox"/> Umwelt <input type="checkbox"/> Umwelt und Klima
1	1	- Neubau von 3 "Doppelhäusern und einem Mehrfamilienhaus (insg. 16 WE) - Neubau einer Tiefgarage mit ca. 40 Stellplätzen - Abbruch eines Bestandsgebäudes (Nr. 59/59a)	tag_name umwe
1	1	- Wahl der Vizepräsidentin bzw. des Vizepräsidenten des Hamburgischen Verfassungsgerichts -	city, url <input type="checkbox"/> (Blank) <input type="checkbox"/> Aachen <input type="checkbox"/> Bayern <input type="checkbox"/> Berlin <input type="checkbox"/> Bielefeld <input type="checkbox"/> Bonn <input type="checkbox"/> Deutschebahn <input type="checkbox"/> Diverse <input type="checkbox"/> Dormagen <input type="checkbox"/> Duisburg <input type="checkbox"/> Düsseldorf <input type="checkbox"/> Essen <input type="checkbox"/> Frankfurt <input type="checkbox"/> Freiburg <input type="checkbox"/> Gelsenkirchen <input type="checkbox"/> Govdata <input checked="" type="checkbox"/> Hamburg <input type="checkbox"/> Herne <input type="checkbox"/> UNO
2	2	- Werbeanlage - - Werbeanlage am Baugerüst - hier: Verlängerung der Aufstellzeit vom 02.10.2014 bis 31.12.2014	
1	1	- Winternotprogramm für Wohnungslose Menschen - Errichtung von Unterkünften bis 250 Menschen in einem Zeitraum von 2017 für 3 Jahre.	
8	8	" Petra III North " Neubau einer Forschungshalle	
1	1	" Roparun " - Staffellauf	
1	1	"Akteneinsicht zum Vorfall an der Ida Ehre Schule - Debatte durch Transparenz verschärfen" - Aktenvorlage gemäß Artikel Verfassung (HV)	
1	1	"Alstercampus - 1. BA" , 4 Neubauten mit Büro- und Wohnnutzung , Erschließung mit Platzfolge Innenhöfe, Sanierung im Bestand Alstertwiete 3: Tiefgarage 1. u. 2. UG	
6	6	"Alstercampus -2.BA" : 3 Neubauten mit Büro- und Wohnnutzung, Neubau Tiefgarage und Umbau Rettungstreppe Hotel (An der Alster 52)	
1	1	"Anodenzelt" - Errichtung einer Lagerhalle	
1	1	"Antisemitismus - erkennen und begegnen" Entwicklung einer Landesstrategie zur Bekämpfung und Prävention von Antisemi Zwischenbericht des Senats zum Ersuchen der Bürgerschaft vom 18. Dezember 2019 "Jüdisches Leben fördern und Antisemi entgegentreten - Einrichtung des Amtes einer beziehungsweise eines Beauftragten für jüdisches Leben und die Bekämpfung Antisemitismus in Hamburg" (Drucksache 21/19335)	
1	1	"Antisemitismus erkennen und begegnen" Entwicklung einer Landesstrategie zur Bekämpfung und Prävention von Antisemi Zwischenbericht des Senats zum Ersuchen der Bürgerschaft vom 18. Dezember 2019 "Jüdisches Leben fördern und Antisemi entgegentreten - Einrichtung des Amtes einer beziehungsweise eines Beauftragten für jüdisches Leben und die Bekämpfung Antisemitismus in Hamburg (Drs. 21/19335)	
1	1	"ARAL" Tankstelle erhält "REWE To Go" Geschäft	
1	1	"Börsentag Hamburg" am 25.10.2014 in den Räumlichkeiten der Handelskammer	
1	1	"Deutsche Allianz Meeresforschung" (DAM) Unterrichtung zu "Deutschen Allianz Meeresforschung e.V." (DAM)	
151.315	248.977		

# Auswertung: Formatangaben der Ressourcen

Die Liste zeigt die Anzahl von Ressourcen mit einem bestimmten Namen und Format mit einem Filter auf Tags mit „umwe“ im Namen.

1. Es gibt offenbar viele Ressourcen mit dem gleichen Namen
2. Es gibt viele Ressourcen ohne Namen
3. Es gibt „Download /INSPIRE“, bei denen nicht klar ist, was dahinter steckt.

# Resources	resource_name	resource_format	url	tag_name
25.443	Genehmigung nach HBauO	PDF	suche	All
24.411	Metadatenbeschreibung aus dem MetaVer	HTML	suche	
19.837	Festsetzungen (Planzeichnung / Verordnung) als PDF Datei	PDF	suche	
15.032	Begründung des Bebauungsplans als PDF Datei	PDF	suche	
13.385	Download WFS Bebauungsplan (XPlanGML)	GML	suche	
10.557		http://publications.europa.eu/resource/authority/file-type/CSV	ckan.i	
7.065	Änderungsbescheid	PDF	suche	
5.924	Sitzungsprotokoll (Onlinekontext)	HTML	suche	
5.029	GeoPortal.rlp	Webanwendung	daten	
5.029	Onlinekarte	Karte	daten	
5.029	WMS Schnittstelle	WMS	daten	
5.019	GeoPortal.rlp	Webanwendung	ckan.i	
5.019	Onlinekarte	Karte	ckan.i	
5.019	WMS Schnittstelle	WMS	ckan.i	
4.871	Originäre Metadaten	HTML	daten	
4.861	Originäre Metadaten	HTML	ckan.i	
4.316	Vorbescheid nach § 63 HBauO	PDF	suche	
4.042	Sitzungsprotokoll (Druckversion)	PDF	suche	
3.614		http://publications.europa.eu/resource/authority/file-type/XML	ckan.i	
3.009		CSV	ckan.i	
2.558		http://publications.europa.eu/resource/authority/file-type/JSON	ckan.i	
2.383		https://www.iana.org/assignments/media-types/text/csv	ckan.i	
2.081	Download (INSPIRE)	Diverse	ckan.i	
2.081	Download (INSPIRE)	Diverse	daten	
481.003				

# Interaktive Auswertung

- Nachdem die aufbereiteten Daten eingelesen sind, lassen sich die gezeigten Auswertungen zur Exploration der Inhalte der Portale mit wenigen einfachen Schritten erstellen.
- Die aufbereiteten Dateien könnten auch in einem der Portale wieder verfügbar gemacht und regelmäßig aktualisiert bereitgestellt werden.
- Je nach den Fragen bzw. Interessen der Auswerter lassen sich Zusammenhänge finden oder Hypothesen prüfen.
- Durch die bereitgestellten URLs der Objekte lassen sie sich auch sehr schnell im Browser aufrufen und prüfen.
- Der aktuelle Stand ist noch WIP und muss weiter bearbeitet werden, um tatsächlich für Nutzer geeignet zu sein.
- Dazu gehört z.B. auch die Unterstützung einer Visualisierung der Portale und der Zahl der Objekte auf einer Karte.