

**OPAL**  
OPEN DATA PORTAL

# Deliverable D1.1

## Anforderungsanalyse

Autoren: Matthias Wauer

Reviewer: Adrian Wilke

Veröffentlichung	Vertraulich
Fälligkeitsdatum	30.09.2017
Fertigstellung	02.02.2018
Arbeitspaket	AP1
Typ	Bericht
Status	Final
Version	1.0

### Kurzfassung:

Dieser Projektbericht beschreibt den Prozess und die Ergebnisse der Anforderungserfassung und Anforderungsanalyse für das Projekt OPAL. Neben einer systematischen Literaturrecherche wurden Anforderungen aus den Ergebnissen von Fragebögen und eines Workshops sowie Erkenntnissen aus der Fachkonzeption und einer Usability-Untersuchung des mCLOUD-Portals gewonnen (vertraulich, weshalb auch dieser Bericht vertraulich zu behandeln ist). Eine Liste von 22 konsolidierten Anforderungen schließt den Bericht ab.

### Schlagworte:

Anforderungserfassung, Anforderungsanalyse, OPAL, Open-Data-Portal

# Inhalt

<b>1 Einführung</b>	<b>1</b>
1.1 Motivation	2
1.2 Vorgehensweise	2
<b>2 Projektberichte</b>	<b>3</b>
2.1 Projektberichte	3
2.1.1 Eigene Projekte	3
2.1.2 Externe Projekte	3
2.2 Fragebogen	5
2.3 Workshops	6
2.4 Bestehende Anforderungen	7
2.5 Eigene Anforderungen	7
<b>3 Anforderungen</b>	<b>7</b>
3.1 Anforderungen aus eigenen und externen Projekten	7
3.1.1 Anforderungen aus LOD2	7
3.1.2 Anforderungen aus LIMBO	8
3.1.3 Anforderungen aus externen Projekten	9
3.1.3.1 Anforderungen aus SmartOpenData	9
3.1.3.2 Anforderungen aus ROUTE-TO-PA	10
3.1.3.3 Anforderungen aus DaPaaS	10
3.1.3.4 Anforderungen aus GovEx-Labs-Dokument	10
3.1.3.5 Weitere relevante Aspekte	10
3.2 Anforderungen aus dem Fragebogen	11
3.2.1 Auswertung des Fragebogens	11
3.2.2 Ableitung der Anforderungen aus dem Fragebogen	17
3.3 Anforderungen aus Workshops	18
3.4 Bestehende Anforderungen aus der Entwicklung von mCLOUD	19
3.5 Eigene Anforderungen	19
<b>4 Zusammenfassung</b>	<b>19</b>
4.1 Konsolidierte Anforderungen	20
4.2 Ausblick	22

### **& : ^cz1j c\**

Ein erfolgreicher Abschluss eines Forschungsprojektes setzt voraus, dass Projektergebnisse wie Prototypen, Softwarekomponenten und im Fall von OPAL das Open-Data-Portal hinsichtlich konkreter Ziele entwickelt werden. Diese Ziele sind im Projektantrag in Grundzügen beschrieben. Zur Umsetzung dieser Ziele ist es erforderlich, dass sie genauer spezifiziert werden. Hierzu werden zunächst Anforderungen erfasst und analysiert.

### **&# Bdi KM ^dc'**

Für die Entwicklung eines Portal-Prototyps und der dafür notwendigen Komponenten ist ein gemeinsames Anwendungsdomänen- und Problemverständnis zwischen den Projektbeteiligten erforderlich.

Daraus ergeben sich die folgenden Ziele der Anforderungsanalyse:

1. Festlegung der Funktionalität des OPAL-Prototyps
2. Überprüfung der Vollständigkeit und Konsistenz dieser Festlegung
3. Erfassung der nicht-funktionalen Eigenschaften von OPAL

### **&# Kdg\Zj Zchl Z'hz'**

Die Erfassung der Anforderungen erfolgt auf der bekannten und verbreiteten Methode auf Grundlage von Anwendungsfällen (Use Cases). Anwendungsfälle beschreiben, wie ein System den Benutzer in seinen fachlichen Aufgaben unterstützt. Aus diesen lassen sich einzelne Anforderungen ableiten. Das Vorgehen kann auf funktionaler Seite in die Erfassung der Ziele (Business Goals), Nutzeranforderungen (User Requirements) und funktionalen Anforderungen (Functional Requirements) unterteilt werden. Auf der anderen Seite stehen die nicht-funktionalen Anforderungen. Diese können sowohl durch Ziele als auch durch Anwendungsfälle gegeben sein. Die gesamtheitliche abschließende Anforderungsspezifikation (Requirements Specification) ergibt sich aus den funktionalen und nicht-funktionalen Anforderungen.

Zur Erfassung der Anforderungen in OPAL wird ein systematischer Ansatz gewählt. Zunächst definieren wir, welche Quellen für die Anforderungserfassung zur Verfügung stehen:

- Eigene Projekte, d.h. aktuelle und abgeschlossene Projekte zu Open-Data-Portalen, an denen wir beteiligt sind bzw. waren
- Externe Projekte, die sich mit Open-Data-Portalen befassen und Anforderungen spezifiziert haben
- Potenzielle und tatsächliche Nutzer von offenen Daten bzw. Datenportalen (insbesondere mCLOUD und MDM)
- Anforderungen, die das BMVI als Betreiber von mCLOUD und MDM selbst erfasst hat

Anschließend legen wir fest, auf welchen Wegen wir diese Anforderungen ermitteln können:

- Sammlung von Anwendungsfällen und Anforderungen aus Projektberichten und -ergebnissen
- Befragungen potenzieller und tatsächliche Nutzer
- Workshops (z.B. der Begleitforschung)
- Anforderungsdokumentation zu relevanten Portalen wie mCLOUD

## D1.1 - Anforderungsanalyse

### Anforderungen der Projektpartner

Diesem Vorgehen entsprechend werden ggf. Anforderungen erfasst, die nicht direkt einem Anwendungsfall zugeordnet werden können. Beispielsweise kann dies vorkommen, wenn relevante Anforderungen aus externen Projekten übernommen werden, die diese nicht über den Ansatz der Anwendungsfälle ermittelt haben, oder sich die Anwendungsfälle nicht zurückverfolgen lassen.

#### ' : g b 1 i j c \ k d c F j Z a z c j c Y K d j M g z 1 j c \ Y z g 6 c [ d j z g c \ h v c v a r z

Entsprechend der Vorgehensweise ermittelten wir geeignete Quellen und Ansätze zur Erfassung der Anforderungen.

#### ' # & E g l z i M g X j i z

Eine gängige Vorgehensweise im Bereich Projektentwicklung ist die initiale oder kontinuierliche Erfassung von Anforderungen. Da sich in der Vergangenheit bereits zahlreiche Projekte mit Open-Data-Portalen beschäftigten oder weiterhin beschäftigen, werden diese als Quellen für die Anforderungsanalyse herangezogen.

#### ' # & : ^ z c z E g l z i z

Einige Projekte des Lehrstuhls Data Science der Universität Paderborn und der Arbeitsgruppe AKSW, aus der dieser Lehrstuhl hervorgegangen ist, befassen sich mit Open-Data-Portalen. Die folgende Tabelle 1 bietet eine Übersicht. In der Spalte "Anforderungen" ist gekennzeichnet, ob die Projekte zugängliche Ergebnisse einer Anforderungsanalyse bereitstellen.

E g l z i	7 z h j g ^ v j c \	A M [ c z 1	6 c [ d j z g c \ z c
LOD2/publicdata.eu	Paneuropäisches Datenportal, aus dem Projekt Linked Open Data 2 (LOD2) hervorgegangen	2010-2014	ja <sup>1</sup>
leipzig-data.de	Initiative zur Veröffentlichung kommunaler Offener Daten	2012-2013	nein
LIMBO	mFUND-Forschungsprojekt zu verknüpften Mobilitätsdaten	2017-2020	ja <sup>2</sup>

I W l a z & Übersicht über eigene Projekte

#### ' # & : n i z g c z E g l z i z

Neben den oben aufgeführten eigenen Projekten existiert eine Vielzahl von externen Projekten zum Thema Open-Data-Portale. Zur Ermittlung dieser wurde eine systematische Literaturrecherche durchgeführt, für die neben einer klassischen Suchmaschine auch Google Scholar sowie Twitter Verwendung fanden. Damit konnten zusätzlich akademische und soziale

<sup>1</sup> Das Anforderungs-Deliverable D1.1 war auf der offiziellen LOD2-Seite zum Zeitpunkt der Erstellung des Berichts nicht verfügbar (<http://static.lod2.eu/Deliverables/D1.1-Requirements.pdf>), konnte aber über das Internet Archive (Wayback Machine) bezogen werden.

<sup>2</sup> Die Anforderungen werden in diesem Projekt parallel zu den Arbeiten in OPAL erfasst, daher können nur vorläufige Ergebnisse z.B. aus einem Fragebogen herangezogen werden.

## D1.1 - Anforderungsanalyse

Netzwerkinformationen zur Ermittlung von Anforderungen verwendet werden. Als Suchbegriffe dienten “Open Data Portal” in Kombination mit “requirements” bzw. “Anforderungen”. Für die generische und die akademische Suchmaschine wurden jeweils die ersten 10 Ergebnisseiten manuell ausgewertet. Aus dieser Datenbasis wurden potenziell relevante Ergebnisse selektiert und anschließend die eigentlichen Dokumente manuell untersucht.

In diesem ersten Schritt wurden insgesamt 77 relevante Ergebnisse (26 Web, 32 akademisch, 19 sozial) gefunden, die in der nachfolgenden manuellen Untersuchung auf 9 tatsächlich relevante Projekte und Dokumente (6 Web, 3 Akademisch, 1 sozial) eingegrenzt werden konnten. Die Liste der daraus ermittelten Ergebnisse ist in Tabelle 2 dargestellt.

Projektname	Beschreibung	URL	Relevanz
ADEQUATE	Untersuchung der Datenqualität auf österreichischen Datenportalen	<a href="https://www.adequate.at/publications-open-material/">https://www.adequate.at/publications-open-material/</a>	ja <sup>3</sup>
SmartOpenData	Verwendung von LOD zum Umweltschutz	<a href="http://www.smartopendata.eu/public-deliverables">http://www.smartopendata.eu/public-deliverables</a>	ja
ROUTE-TO-PA	Soziale Plattform für Offene Daten, Portalerweiterungen für verbesserte Transparenz	<a href="http://routetopa.eu/">http://routetopa.eu/</a>	ja
DaPaaS	Offene Plattform zur Veröffentlichung von Open Data als Cloud-Service	<a href="https://project.dapaas.eu/">https://project.dapaas.eu/</a>	ja <sup>4</sup>
GovEx Labs (Center for Government Excellence)	“Open Data Portal Requirements” - kollaborative Sammlung von typischen Anforderungen	<a href="http://labs.centerforgov.org/open-data/portal-requirements/">http://labs.centerforgov.org/open-data/portal-requirements/</a>	ja
Open Data Handbook	Anleitung zur Erstellung und Bereitstellung offener Daten	<a href="http://opendatahandbook.org/guide/en/how-to-open-up-data/">http://opendatahandbook.org/guide/en/how-to-open-up-data/</a>	ja <sup>5</sup>
Leitfaden zum Open Data Portal Bayern	Hinweise zur Bereitstellung offener Daten	<a href="https://opendata.bayern.de/download/opendata-leitfaden-ver-637C2FA80DCF6EDEF741D29E98468F4.pdf">https://opendata.bayern.de/download/opendata-leitfaden-ver-637C2FA80DCF6EDEF741D29E98468F4.pdf</a>	ja <sup>6</sup>

<sup>3</sup> Relevantes Deliverable D1.2 derzeit nicht öffentlich verfügbar

<sup>4</sup> sehr eingeschränkte funktionale Anforderungen in Deliverable 1.1

<sup>5</sup> eher sinnvolle Best Practices

<sup>6</sup> eher sinnvolle Best Practices

## D1.1 - Anforderungsanalyse

Evaluierung von GovData	Ergebnisbericht zur Evaluierung des GovData-Portals	<a href="https://www.govdata.de/documents/10156/18448/Evaluierung_GovData_Evaluationsbericht_final.pdf/b89d3681-b8c1-4864-a8f5-ed2483c1a5d5">https://www.govdata.de/documents/10156/18448/Evaluierung_GovData_Evaluationsbericht_final.pdf/b89d3681-b8c1-4864-a8f5-ed2483c1a5d5</a>	nein <sup>7</sup>
Workshop zum Berliner Open Data Portal	Dokumentation von Ablauf und Fazit eines Open-Data-Portal-Workshops	<a href="https://projektzukunft.berlin.de/fileadmin/user_upload/pdf/IKT-Wirtschaft/Datenportal_2_0_WS4_BODDy.pdf">https://projektzukunft.berlin.de/fileadmin/user_upload/pdf/IKT-Wirtschaft/Datenportal_2_0_WS4_BODDy.pdf</a>	nein

### I WAZ<sup>9</sup> / Externe Projekte und Quellen

#### ' # ; gAZVZc'

Zur weiteren Ermittlung von Anforderungen durch potenzielle und tatsächliche Nutzer solcher Portale wurde ein Fragebogen entwickelt (sowohl in Deutsch als auch in Englisch) und an wichtige Mailinglisten und Verteiler wie OKFN-DE gesendet. Die folgende Tabelle 3 erläutert den Aufbau des Fragebogens.

; gAZ	6ci dj[db V	: g VjZg>ej i
Nutzen Sie offene Daten? Wenn ja, welche? Für welche Anwendung?	Freitext	Kurze Beschreibung von genutzten Datenquellen und derzeitigen Anwendungen
Auf welchem Weg finden Sie derzeit diese Daten?	Mehrfachauswahl mit Ergänzungsoption	Derzeitige bevorzugte Möglichkeiten bei der Suche nach offenen Daten
Wie bewerten Sie die derzeitige Möglichkeit, offene Daten zu finden und zu nutzen?	Lineare Likert-Skala (1-5)	Generelle Zufriedenheit mit derzeitigen Möglichkeiten
Welche Schwierigkeiten sehen Sie in den derzeitigen Möglichkeiten, offene Daten zu finden und zu nutzen?	Freitext	Kurze Beschreibung konkreter Schwachpunkte und Problemstellen derzeit verfügbarer Möglichkeiten
Welche Verbesserungsvorschläge haben Sie für diese Schwierigkeiten?	Freitext	Positives Nutzerfeedback zu konkreten Ansätzen zur Optimierung von Portalen
Welche Metadaten (Eigenschaften) von offenen Datensätzen sind aus Ihrer Sicht für eine erfolgreiche Nutzung entscheidend?	Mehrfachauswahl mit Ergänzungsoption	Hinweise auf Metadaten, die für Nutzer besonders relevant sind und auf die sich die Optimierung ggf. konzentrieren sollte
Wie wichtig ist für Sie die Kombination (Integration) mehrerer Datenquellen?	Lineare Likert-Skala (1-5)	Subjektive Einschätzung der Wichtigkeit von AP4 und 5

<sup>7</sup> vor allem Feststellungen aus der Untersuchung, aus denen sich ggf. Anforderungen ableiten lassen

## D1.1 - Anforderungsanalyse

Wie hilfreich fänden Sie eine mobile Anwendung zur Suche nach offenen Daten an Ihrem Standort?	Lineare Likert-Skala (1-5)	Subjektive Einschätzung der Wichtigkeit des "Mobile App"-Prototyps
Wie hilfreich fänden Sie einen Assistenten, der geeignete offene Daten zu natürlichsprachlichen Anfragen (z.B. "Hat jemand die Geokoordinaten von allen Bahnübergängen in Deutschland?") vorschlägt?	Lineare Likert-Skala (1-5)	Subjektive Einschätzung der Wichtigkeit des "Social Bot"-Prototyps
Haben Sie weitere Anmerkungen?	Freitext	Zusätzliche Hinweise, die nicht zuvor angegeben werden konnten

I WZL ( / ; gAZVtZc

' # L dg h| deh

Im Rahmen des Kickoff-Workshops der mFUND-Begleitforschung am 7. Dezember 2017 in Bonn wurden in der Arbeitsgruppe 2 durch die Teilnehmer Anforderungen an das Datenportal mCLOUD erfasst. Diese umfassten sowohl Anforderungen an mCLOUD direkt, aber auch an die Datenanbieter und Dritte (z.B. Gesetzgeber). Abbildung 1 zeigt die Ergebnisse der Anforderungserfassung.





Ergebnisse der Anforderungserfassung der mFUND-Begleitforschung

Die Ergebnisse und konkrete Anforderungen werden in Abschnitt 3 diskutiert.

### 7.2.1.2 Ergebnisse der Anforderungserfassung

Da mit mCLOUD bereits ein Portal für offene Daten entwickelt wurde, untersuchten wir, welche Anforderungen dabei bereits erfasst und ggf. evaluiert wurden. Über den dafür zuständigen Mitarbeiter beim BMVI erhielten wir zwei Dokumente, ein Fachkonzept zum Datenportal und die Ergebnisse einer Usability-Studie. Beide Dokumente sind vertraulich, sodass wir dieses Deliverable entsprechend ebenfalls als vertraulich deklarieren, um entsprechende Informationen einarbeiten zu können.

### 7.2.1.3 Anforderungen an die mCLOUD

Zusätzlich soll die Definition eigener Anforderungen, die aus unserer Erfahrung mit der Nutzung von Open-Data-Portalen im Allgemeinen und mCLOUD bzw. MDM im Besonderen oder aus der Vorhabensbeschreibung entstammen, den Katalog ergänzen. Hinzu kommen Anforderungen, die aus den konkreten Anwendungsszenarien mit den jeweiligen Prototypen in OPAL entstehen.



( : g\ZVt^hZ^YZg^6c[dj^Zg c\hZg^Vhj c\

(#& 6c[dj^Zg c\Zc^Vj hZ^ZcZc^j c^ZriZgcZc^Egd^Z^ iZc^

(#&# 6c[dj^Zg c\Zc^Vj h^AD9^

In Tabelle 4 sind die Anforderungen aus dem entsprechenden LOD2/publicdata.eu-Deliverable zusammengefasst, die im Kontext von OPAL relevant sind. Da in diesem Projekt ein “community-driven data portal” entstand, weichen die Anforderungen z.T. deutlich von denen eines Metadatenportals wie OPAL ab. Außerdem sind vor allem der “OGD Use Case” (Kapitel 4) und die entsprechenden Anforderungen in den “consolidated requirements” (Kapitel 5) für OPAL von Interesse. In der Spalte “Art” steht “F” für funktionale und “NF” für nichtfunktionale Anforderungen.

Anforderung	Art	Beschreibung
AE1	F	Sammeln von Daten (REQ03: Consume/harvest data). Entsprechende Daten können entweder Informationen zum Anreichern der bestehenden Daten, aber auch Datensätze als Metadaten zur Beschreibung von Datensätzen sein.
AE2	F	Speicherung von Daten (REQ06: Store acquired data). Mechanismus zur persistenten Speicherung von Daten bzw. Metadaten, einschließlich der Versionierung von Änderungen.
AE3	F	Integration (REQ07: Integrate data). Mechanismus zur Konvertierung in verschiedene Formate, Mapping auf verschiedene Metadaten-Schemata.
AE4	F	Anzeige (REQ08: Display data). Zeigen von existierenden und neuen Daten im Repository, einschließlich verfügbarer Metadaten.
AE5	F	Analyse (REQ09: Analyse data). Mechanismen zur Analyse neu hinzugefügter Daten und deren Werte (im Rahmen von OPAL v.a. Metadaten): z.B. Inkonsistenzen, Validierung, Syntaxfehler.
AE6	F	Suche (REQ11: Search for data). Fortgeschrittene Suchmechanismen für Metadaten und Datensätze (moderierte Suche, facettierte Suche etc.)
AE7	F	Vorschläge (REQ12: Recommend data). Empfehlungen von relevanten Datensätzen basierend auf semantischer Dokumentenanalyse, Suchanfrage, gewählten Datensätze und persönlichem Profil.
AE8	F	Kommentierung (REQ14: Comment data). Kommentierung bezüglich Qualitäts- und inhaltsbezogener Kriterien.
AE9	F	Bewertung (REQ15: Rate data). Möglichkeit zur Bewertung der Datensätze anhand von Qualitätskriterien.
AE10	F	Verknüpfung (REQ17: Link/Align data). Mechanismen zur Erkennung und Erstellung semantisch sinnvoller Beziehungen zwischen Datensätzen.

## D1.1 - Anforderungsanalyse

AE11	F	Zugriff (REQ19: Access data). Mechanismen für Zugriff und Download der Datensätze; SPARQL, API und Download.
AE12	F	Qualitätsbewertung (REQ21: Quality assessment). Mechanismen zur Prüfung der semantischen Konsistenz, Validität und Repräsentationsqualität (bei OPAL insbesondere der Metadaten).
AE13	F	Überwachung (REQ22: LOD monitoring). Monitoring von Änderungen an Datensätzen (in OPAL insbesondere Verfügbarkeit).

(WAZ): Anforderungen aus dem eigenen Projekt LOD2

### (#) 6c[dyZj c\ZcVj hA:B 7D

Da OPAL mit LIMBO zusammenarbeitet und dies ebenfalls das Ziel verfolgt, die Auffindbarkeit geeigneter offener Daten zu verbessern, wurden die vorläufigen Anforderungen, die bis zum aktuellen Zeitpunkt in LIMBO erfasst wurden, untersucht.

Insgesamt wurden 8 Anwendungsfälle erfasst, von denen für OPAL vor Allem der Use Case "Suchmaschine für Mobilitätsdaten" relevant ist. Konkrete Anforderungen waren jedoch nur für einen anderen Anwendungsfall verfügbar. Für die Suchmaschine hingegen liegen Ergebnisse aus einem Fragebogen vor, der untersuchte, welche offenen Mobilitätsdaten wie genutzt werden und welche Funktionalität eine entsprechende Suche bieten muss.

Für OPAL sind hierbei die erfassten Informationsbedürfnisse und entsprechende Schlüsselwörter relevant, die in Tabelle 5 aufgelistet sind. Zusätzlich wurde untersucht, ob entsprechende Daten in mCLOUD oder auf externen Quellen verfügbar sind. Mit "ja\*" bewertete externe Verfügbarkeit bedeutet, dass die Daten nur teilweise und/oder unter anderen Schlüsselwörtern gefunden werden können. Für mCLOUD müssen ebenfalls häufig andere Schlüsselwörter genutzt werden.

Informationsbedürfnis	Schlüsselwörter	Verfügbar in mCLOUD	Extern Verfügbarkeit
Welche abgasrelevanten Daten von Berlin gibt es?	Abgasdaten Berlin	ja	ja
Gib mir Datensätze für aktuelle Baustelleninformationen?	Baustellen Datensatz	ja	ja
Gib mir Datensätze für Bus- und Zugverspätungen? [sic]	Bus/Zug Verspätung	nein	ja*
GTFS daten für Deutschland?	GTFS Deutschland	ja	ja
Gib mir Echtzeitfahrradzählungsdaten?	Echtzeit Fahrradzählung	ja	nein
Welche offenen Parkdaten von Deutschland gibt es?	Offene Parkdaten Deutschland	ja	ja
Gib mir VBB Transitdaten.	VBB Transitdaten	ja	ja*

## D1.1 - Anforderungsanalyse

Gib mir offene Verkehrsdaten für Berlin.	Offene Verkehrsdaten Berlin	ja	ja
Gibt es Geonames Datensätze?	geonames	nein	ja
Gib mir Historische Wetterdaten für Frankfurt.	historische Wetterdaten Frankfurt	nein	ja*
Gib mir gtfs-Echtzeitdaten für Deutschland?	gtfs echtzeit Deutschland	nein	ja*
Gib mir Echtzeitautozählungen?	Echtzeit Autozählung	nein	nein

I Wiaz\*: In LIMBO erfasste Informationsbedürfnisse

(#&# 6c[dyZg c\ZcVj hZniZgcZcEglZ iZc

(#&#& 6c[dyZg c\ZcVj hi-b VjDeZc9M V

In den Anforderungs-Deliverables des Projekts SmartOpenData werden 5 Anwendungsszenarien definiert, denen anschließend (in D2.1 Abschnitt 3.4) funktionale Anforderungen zugeordnet werden. Diese sind relativ generisch. Schließlich werden in Abschnitt 8 die konsolidierten Anforderungen zusammengefasst. Für OPAL sind nur sehr wenige Anforderungen daraus relevant, die bereits durch eigene Anforderungen (siehe 3.5) hinsichtlich der OPAL-Vorhabensbeschreibung abgedeckt werden. Alle anderen Anforderungen, wie z.B. die Unterstützung privater Daten und Kartenbilder, sind für OPAL nicht erforderlich. Weitere "Service Requirements" definieren in erster Linie Funktionsblöcke (wie Monitoring und Transformation) sowie Formate (wie KML), die das System unterstützen soll. Schließlich werden in Abschnitt 8.3 "spezifische Anforderungen" an die SmartOpenData-Infrastruktur gestellt, die jedoch eher eine Sammlung von Zielbeschreibungen des Systems und Aussagen über die verfolgten Prinzipien und Initiativen darstellt. Sie sind generell zu wenig konkret, um sinnvolle Anforderungen in OPAL darzustellen. Lediglich die folgenden Aspekte können für die Umsetzung von OPAL von Interesse sein:

Unterstützung der relevanten Aspekte von GeoSPARQL, CKAN und DCAT-AP

Nennung geeigneter Vokabulare, z.B. INSPIRE Registry, EU Metadata Registry, EuroVoc und ISA Core Location Vocabulary

Lizenzierung vorrangig unter Creative Commons CC0

Im zweiten Dokument aus SmartOpenData, D2.2 zu Nutzeranforderungen und Anwendungsfällen, werden die 5 Anwendungsfälle ("Pilots") definiert einschließlich der typischen Abläufe. Es listet insbesondere auf, welche Datensätze für diese Anwendungsfälle in welchem Zustand zur Verfügung stehen, und welche Komponenten der SmartOpenData-Plattform benötigt werden. Daher bietet das Dokument aus Sicht von OPAL keine weiteren relevanten Anforderungen.

(#&# 6c[dyZg c\ZcVj hGDJI: 'ID'E6

In Deliverable 2.4 des Projekts werden Anforderungen hinsichtlich zwei Projektzielen definiert. Während das erste Ziel "Social Platform for Open Data" nicht für OPAL relevant ist, sind die User Stories für das zweite Ziel "Transparency Enhancing Toolset" (TET) in dem Sinn relevant, als dass dadurch eine höhere Datentransparenz und -qualität erreicht werden soll. Die User Stories wurden zuvor in Workshops anhand von Anwendungsfällen erfasst.

## D1.1 - Anforderungsanalyse

Leider wird in der abschließenden Anforderungsanalyse nicht mehr nach diesen beiden Projektzielen unterschieden. Eine manuelle Prüfung der TET-Anwendungsfälle ergab, dass die meisten Anforderungen grundsätzlich für OPAL relevant sind, allerdings in TET sehr spezifisch auf CKAN zugeschnitten. Weiterhin sind die Anforderungen nur für die Bearbeitung einer CKAN-Datenquelle gedacht, während OPAL einen viel weiteren Umfang an Funktionalität abdecken soll.

### ~~(##)~~ 6c[dy]Zg c\Zc\j h<9VEWH

Im DaPaaS-Projektbericht D1.1 sind lediglich generelle Anforderungen von vorab festgelegten Anwenderklassen der Plattform erfasst worden. Aus Sicht von OPAL sind lediglich die Anforderungen der Nutzerrollen "Application Developer" und "End-user Data Consumer" relevant (Abschnitte 2.2.3 und 2.2.4). Diese beschränken sich auf die Verfügbarkeit einer API, Datenexport, Suchfunktion, Lesezeichen und der nicht-funktionalen Anforderung hoher Verfügbarkeit.

### ~~(##)~~ 6c[dy]Zg c\Zc\j h<dk r^AMM'9d j bZci

In "Open Data Portal Requirements" hat GovEx Labs einen kollaborativen Ansatz gewählt, Anforderungen für ein Open-Data-Portal aus Sicht von 3 Nutzerrollen (Konsument, Anbieter und Administrator) zusammenzutragen. Das Dokument enthält insgesamt 41 Anforderungen, davon 15 für die aus OPAL-Sicht relevante Nutzerrolle des "Consumers".

Insgesamt sind die Anforderungen relativ generisch und umfassen z.B. die Suche nach einem oder mehreren Begriffen, die explorative Suche nach Kategorien, eine Zusammenfassungsseite oder Visualisierungen. Diese sind für OPAL relevant, allerdings auch sehr naheliegend.

### ~~(##)~~ L Z1Zg7gZaK\ciZ'6heZ iZ

In diesem Abschnitt werden weitere relevante Informationen aus externen Quellen zusammengefasst, die im Rahmen von OPAL interessant sind, aber nicht unmittelbar Anforderungen nach sich ziehen.

Für die Weiterentwicklung des Open-Government-Data-Modells (OGD) für das deutsche Portal GovData wurden 12 Bereitsteller von Daten nach den von ihnen meistgenutzten Metadaten befragt. Abbildung 2 fasst die Ergebnisse zusammen. Im Hinblick auf OPAL ist dies insbesondere relevant, weil eine unterschiedliche Bewertung der Wichtigkeit eines Metadatum durch Anbieter und Nutzer zu dem Problem führen kann, dass notwendige Angaben fehlen oder hoher Aufwand in die Dokumentation von ggf. unnötigen Metadaten investiert wird.

## D1.1 - Anforderungsanalyse

Nr	Metadaten-Element in OGD 1.1	Häufigkeit der Nutzung
1	Titel	10
2	Beschreibung	10
3	Kategorien	10
4	Ressourcen	10
5	URL	10
6	Format	10
7	Lizenz-ID	9
8	Name	8
9	Veröffentlichende Stelle	8
10	Datenverantwortliche Stelle	8
11	Website	8
12	Typ	8
13	Beschreibung	8
14	Veröffentlichende Stelle Email	7
15	Datenverantwortliche Stelle Email	7
16	Sprache	7
17	Kalender-Daten	7
18	Rolle	7
19	Datum	7
20	Start-Datum	7
21	End-Datum	7

Nr	Metadaten-Element in OGD 1.1	Häufigkeit der Nutzung
40	Adresse	2
41	Namensnennungs-Text	2
42	Nutzungsfreiheit	2
43	Original-Metadaten-Portal	2
44	Räumliche Abdeckung der Datensätzen	2
45	Geographische Abdeckung durch den Amtlichen Gemeindegeschlüssel (AGS).	2
46	Geographische Abdeckung durch NUTS-Code	2
47	Bilder	2
48	Freitext	1
49	Unterkategorien	1
50	Original-Metadaten-XML	1
51	Metadaten-Transformator	1
52	Abdeckung in Koordinaten	1
53	Verwendete Datensätze	1
54	Original-Metadaten-HTML	0
55	Linked Data Ressource die per URI identifiziert ist und der geographischen Abdeckung entspricht.	0
56	Abdeckung als Freitext falls keine der Nomenklaturen oder Standards verwendet werden können.	0

**Abbildung 1**: Häufigkeit der Nutzung von Metadatenelementen von Datenanbietern auf GovData<sup>8</sup>

Im COMSODE-Projekt<sup>9</sup> wurde versucht, mit dem Open Data Node ein Werkzeug und eine Methodik zur vereinfachten Bereitstellung und Nutzung von offenen Daten zu entwickeln. Im Deliverable D2.1<sup>10</sup> sind in Abschnitt 4 Anforderungen an eine solche „Publikationsplattform“ erläutert. Der Schwerpunkt liegt jedoch offenbar auf der Bereitstellung von Daten (über das ETL-Werkzeug UnifiedViews), die Plattform-Funktionalität für Nutzer scheint eine Erweiterung von CKAN hinsichtlich Visualisierung und API-Generierung zu sein.

**2.1.1** Anforderungen an die Plattform

**2.1.1.1** Anforderungen an die Plattform

Insgesamt erhielten wir 17 verwertbare Antworten für den Fragebogen (siehe 2.2). Im Folgenden werden zunächst die Antworten zum Ist-Zustand untersucht, bevor Anforderungen aus dem von den Teilnehmern beschriebenen Soll-Zustand abgeleitet werden.

Alle 17 Teilnehmer nutzen bereits offene Daten. Zentrale Themen sind Verkehrsdaten, die z.B. für Netzwerk- und Erreichbarkeitsanalysen sowie Fahrpläne genutzt werden, sowie kommunal- und entwicklungspolitische Zwecke (siehe Tabelle 6).

Übersicht über die Metadaten-Elemente in OGD 1.1
Übersicht über die Häufigkeit der Nutzung
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform
Übersicht über die Anforderungen an die Plattform

<sup>8</sup> <http://www.digitaler-staat.org/wp-content/uploads/2017/02/Horn.pdf>

<sup>9</sup> <http://comsode.eu>

<sup>10</sup> [http://www.comsode.eu/wp-content/uploads/COMSODE\\_D2.1\\_Final.pdf](http://www.comsode.eu/wp-content/uploads/COMSODE_D2.1_Final.pdf)

## D1.1 - Anforderungsanalyse

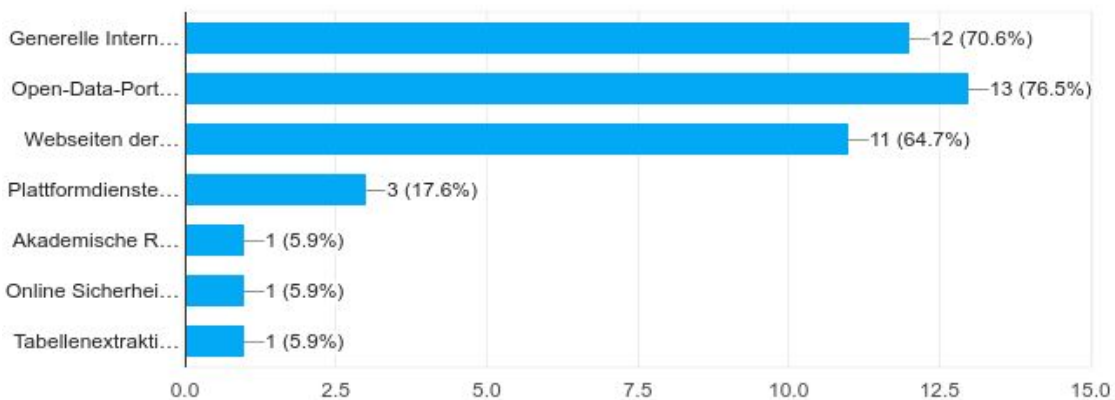
Reiz: /AZ^a@e] • e&@&Ucaicā {i{ aca }^} Ksa ā^i* e:ē) EÜ] a    éc ^EÜ&@  ^} Á ~^} ^i/ Pæ • @pÁ ^c&EÄ
Ö> ÁaÁÜ aē •^Á} áÁ[  i^ aca } Á aÁē) á^i^} /Öae } Á
Q   •&@ } * EÖ^ } [ { ^aē aē •^Á
Reiz: [  Áē^ { Á^i\^@ • EÄ } áÁ aē^i\^@ • áae } Á Ç EÖEÖae@   e } ^DÁ
T ^ā c } • Á>  Áā ^} ^ÁY ^áÁÜ: [ b\ c EÜ]  ā ^ÁÜ&@  &@ āÁ } áÁ aē ā aca } EÄ
Ü~ &@Á
Ö) ç a\ c Áā ā ^ÁÖ:ē  i^Á>  Áae Áā: ā ^iÁÜ ] ^} ÖaeÁÜ [  caE * ^* a\ c ÁÜXÁ } Á ] ~  áÁ aē E EÜ [  caÁ>  Á ā •^} •&@ca&@ Á, ^&^ EÄ
ÖÖ] ^āā EÖPÖÁ>  Á ~  cā ^āāÁÜ } [ caē } Á } áÁÜ^ { aē cÁÜ^ aē&@Á
Y ā •^} •&@ca&@ Á, ^&^ EÄ ^āā, ^} á~ } *^} Á
T [ áāāe o áae } Áē Á ÖŠUWÖÁ
Öà] ^āāÁÁā \^áÖ^ [ ÖaeÁÜ aē  áae ^ÁÜ aē •^} Á

**IWIWZ**+: Antworten zur derzeitigen Nutzung von offenen Daten

Die offenen Daten werden dabei vor allem über Open-Data-Portale (77%), generelle Internetsuchdienste (71%) und Webseiten der Datenanbieter (65%) gefunden (Abbildung 3).

Auf welchem Weg finden Sie derzeit diese Daten? Bitte geben Sie wenn möglich konkrete Quellen im letzten Feld an.

17 responses

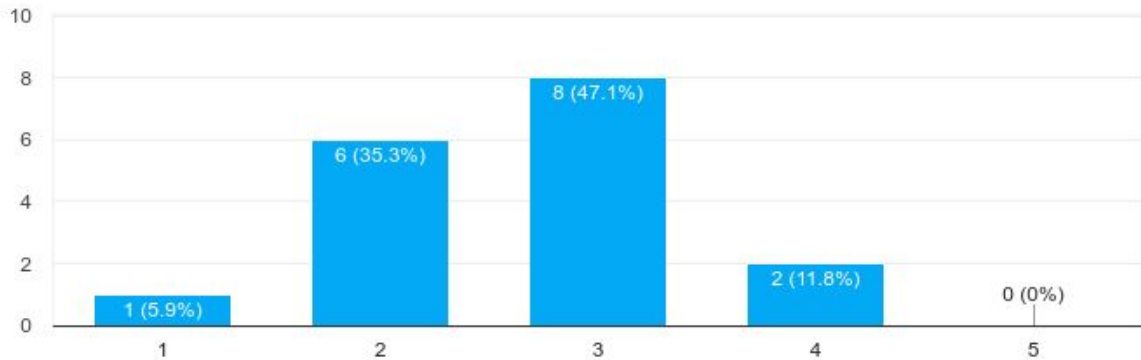


**ÖWÖY c\`:** Quellen zur Suche nach offenen Daten

Keiner der Teilnehmer ist vollständig zufrieden mit den derzeitigen Möglichkeiten, offene Daten zu finden (Abbildung 4).

## Wie bewerten Sie die derzeitige Möglichkeit, offene Daten zu finden und zu nutzen?

17 responses



**6WMBj c\**): Bewertung der Suche nach offenen Daten (Mittelwert: 2,65; Standardabweichung: 0,79)

Die Schwierigkeiten, die zu dieser Bewertung führen, sehen die Teilnehmer vor Allem in inkonsistenten Daten / Datenformaten und generellen Qualitätsproblemen, schlechter Verfügbarkeit (z.B. 404-Fehler) und mangelnder Zugriffsoptionen (z.B. fehlende Datendumps oder problematische APIs), sowie fehlende Vernetzung und Aggregation von Datensätzen (jeweils 4 Nennungen). Weitere häufige Probleme sind eine schlechte Auffindbarkeit von Daten und -portalen sowie geringe Auswahl an relevanten Daten (jeweils 3 Nennungen), gefolgt von fehlender Offenheit der Daten, fehlender Metadaten und dem Problem, dass zu viele einzelne Portale existieren (jeweils 2 Nennungen, Abbildung 5).



### Häufige Probleme

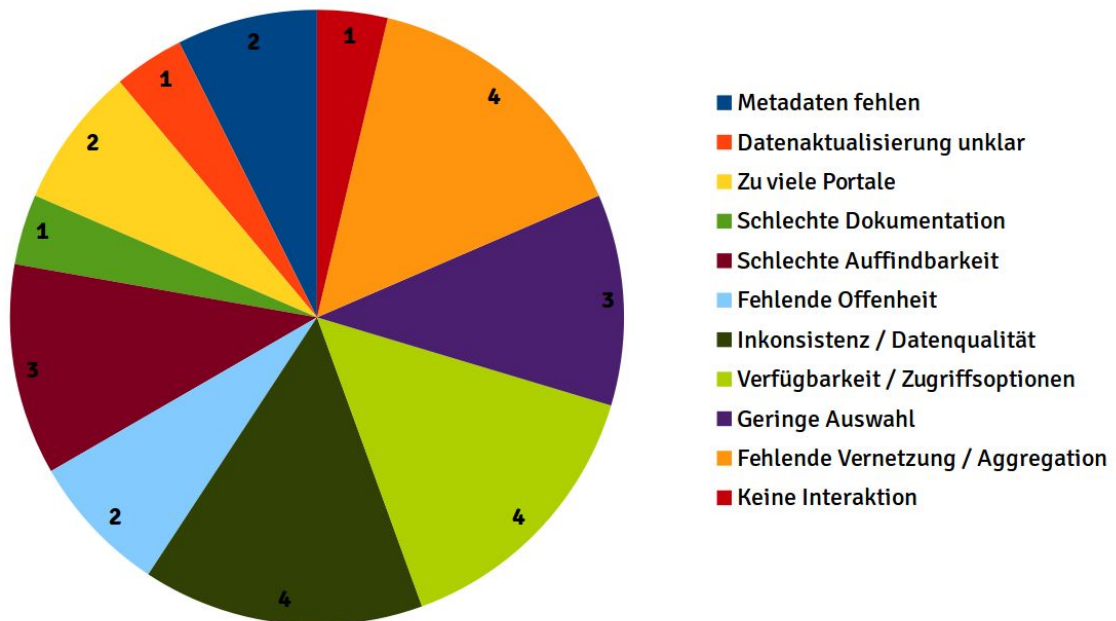


Abbildung 6: Typische Schwierigkeiten bei der Nutzung offener Daten

Die Teilnehmer nannten zahlreiche Vorschläge zur Verbesserung (Abbildung 6). Häufig genannt wurden besser aufbereitete Daten (insbesondere Linked Data und die klare Auszeichnung vollständiger und korrekter Metadaten und Lizenzangaben). Weitere Aspekte waren die Nutzung einheitlicher Datenformate und die Konzentration auf ein zentrales Datenportal sowie Standard-APIs anstelle von proprietärer Lösungen. Weiterhin sind eine optimierte Suche, die kontinuierliche Überprüfung der Verfügbarkeit von Datensätzen und die Annotation der Datensätze auf Webseiten mit z.B. maschinenlesbaren Microformaten genannt worden. Weniger das Datenportal als vielmehr die Datenbereitsteller und Dritte werden durch die Empfehlung besserer Dokumentation (besonders der Erstellungsprozesse), Mehrsprachigkeit, eine automatisierte (ggf. kontinuierliche) Datenbereitstellung und die Durchsetzung weltweit einheitlicher Lizenzen (im Gegensatz zu für Deutschland spezifischen Lizenzen) adressiert.

### Verbesserungs-/Lösungsvorschläge

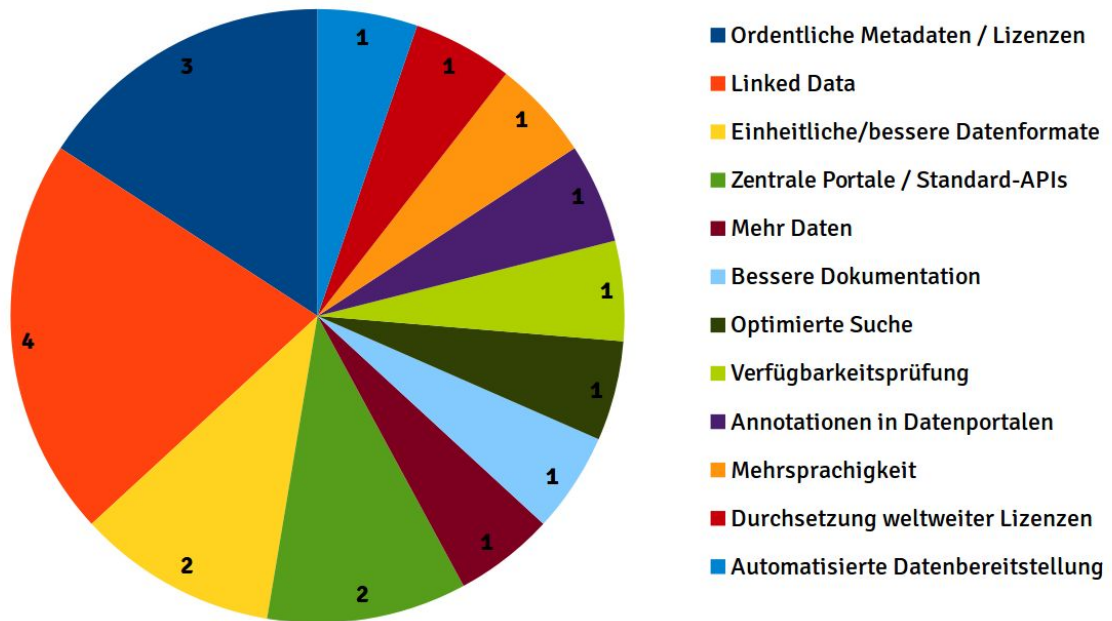


Abbildung 6: Verbesserungs-/Lösungsvorschläge.

Welche Metadaten für die Nutzer besonders wichtig sind, zeigt Abbildung 7. Für den Großteil der Nutzer sind der Inhalt, die Lizenz und das Format entscheidend. Ebenfalls sehr relevant sind die Qualität und das Aktualisierungsintervall der Daten, gefolgt von Umfang, Zeitraum, räumlicher Ausdehnung und dem Zugriffsprotokoll.

### Welche Metadaten (Eigenschaften) von offenen Datensätzen sind aus Ihrer Sicht für eine erfolgreiche Nutzung entscheidend?

17 responses

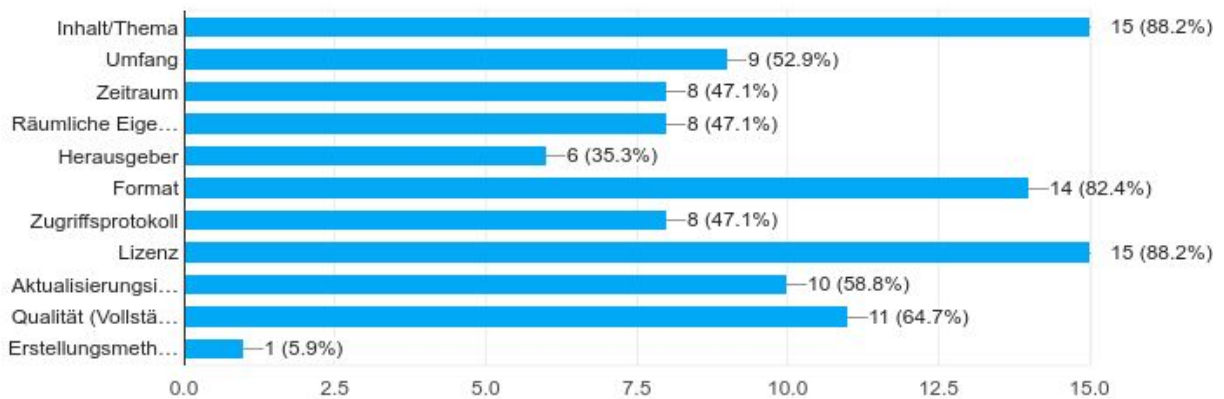
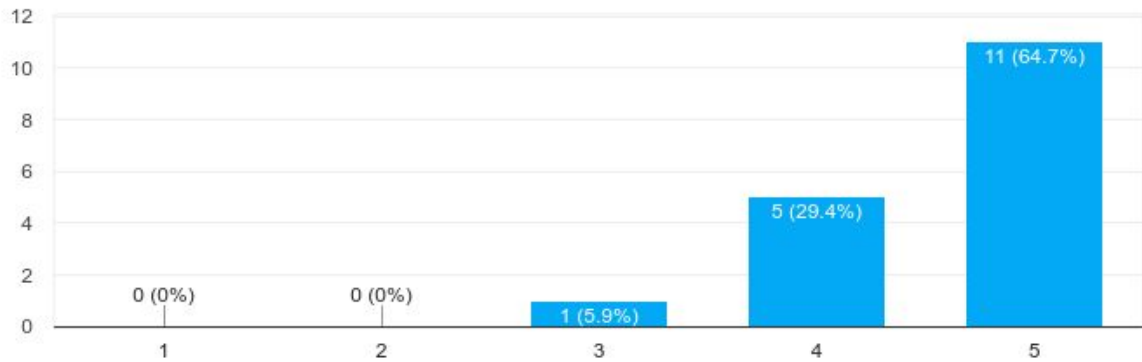


Abbildung 7: Wichtige Metadaten

Für fast alle Teilnehmer ist die Integration von Datensätzen von sehr hoher Bedeutung (Abbildung 8). Hingegen sind mobile Anwendungen und soziale Assistenten zum Finden offener Daten für einen Teil relevant (Abbildungen 9 und 10).

### Wie wichtig ist für Sie die Kombination (Integration) mehrerer Datenquellen?

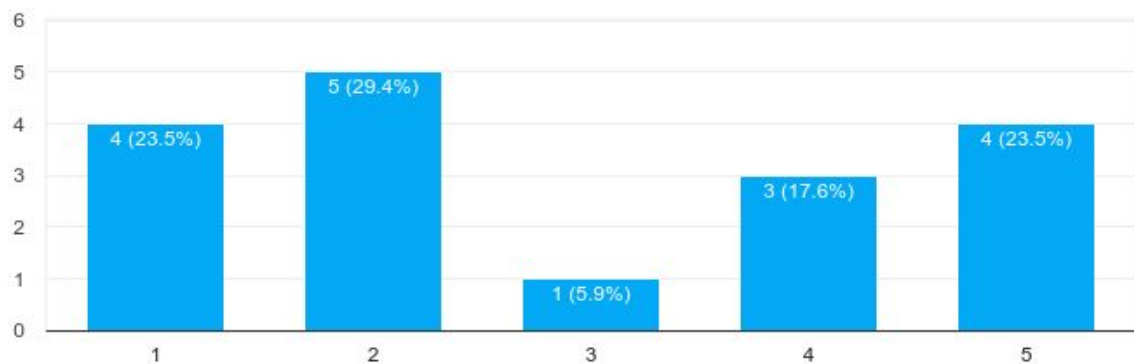
17 responses



**SWBj c\** : Wichtigkeit der Integration (Mittelwert: 4,6; Standardabweichung: 0,6)

### Wie hilfreich fänden Sie eine mobile Anwendung zur Suche nach offenen Daten an Ihrem Standort?

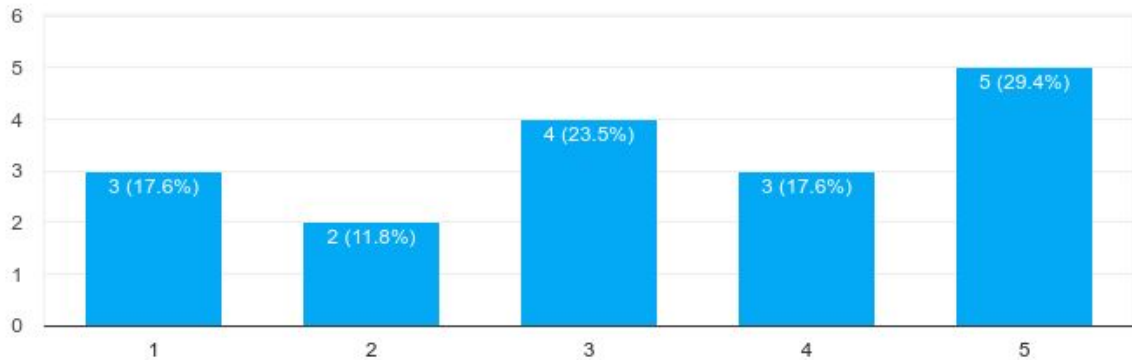
17 responses



**SWBj c\** : Nützlichkeit einer mobilen Suche (Mittelwert: 2,9; Standardabweichung: 1,6)

Wie hilfreich fänden Sie einen Assistenten, der geeignete offene Daten zu natürlichsprachlichen Anfragen (z.B. "...rgängen in Deutschland?") vorschlägt?

17 responses



60% Nützlichkeit eines sozialen Assistenten (Mittelwert: 3,3; Standardabweichung: 1,5)

(# # 60% Nützlichkeit eines sozialen Assistenten (Mittelwert: 3,3; Standardabweichung: 1,5)

Anforderung	Art	Beschreibung
AF1	F	Entwicklung eines komponentenbasierten Meta-Datenportals mit umfangreichen Suchmöglichkeiten. Hintergrund ist, dass die Teilnehmer offene Daten meist über Datenportale suchen, aber gleichzeitig bemängeln, dass es zu viele verschiedene Portale gibt. Daher sollen möglichst wiederverwendbare Komponenten entwickelt werden, die ggf. in künftigen zentralen Portalentwicklungen wiederverwendet werden können.
AF2	F	Programmatischer Zugriff über standardisierte API. Wenn möglich sollen mehrere Zugriffsmethoden und Datenformate bereitgestellt werden, um den unterschiedlichen Nutzeranforderungen zu genügen.
AF3	F	Konvertierung der Metadaten in Linked Data. Ziel ist die Repräsentation in einem einheitlichen Format, das eine bessere Verknüpfung und Integration ermöglicht.
AF4	F	Analyse der (Meta-)Datenqualität. Um Inkonsistenzen und unzureichende Informationen zu erkennen muss die Qualität dieser Daten mit geeigneten Metriken untersucht werden. Eventuell muss hierzu Nutzerfeedback eingebunden werden.
AF5	F	Kontinuierliche Prüfung der Verfügbarkeit. Durch den Crawling-Ansatz, den OPAL verfolgt, ist ein kontinuierliches Recrawling der Datensätze möglich (AF4.1). Um die Verfügbarkeit der eigentlichen Datensätze sicherzustellen sollen für die meisten Zugriffsprotokolle geeignete Konnektoren bereitgestellt werden, die den Zugriff auf die Daten überprüfen (AF4.2) und durch geeignete Eigenschaften (z.B.

## D1.1 - Anforderungsanalyse

		Änderungsdatum bei FTP, E-Tag bei HTTP) auch das Aktualisierungsintervall ableiten können (AF4.3).
AF6	F	Automatisierte Integration von (Meta-)Daten. Als Ansatz gegen eine fehlende Verknüpfung und Aggregation von Daten, hohe Relevanz auch durch die Rückmeldungen in Abbildung 8.
AF7	F	Prüfung und Integration von Lizenzinformationen. Motivation ist die hohe Bedeutung korrekter Lizenzinformationen, als auch die Notwendigkeit, resultierende Lizenzbedingungen aus der Integration von mehreren Datensätzen abzuleiten.
AF8	NF	Qualität der Suchfunktionalität. Gegen das Problem der schlechten Auffindbarkeit soll durch geeignete Maßnahmen die Vollständigkeit und Präzision der Suchergebnisse, insbesondere bei Suchanfragen die andere Begriffe als die Datensatzbeschreibung verwenden, verbessert werden.

I **WiaZ**<sup>°</sup>: Anforderungen aus Fragebögen

### (# 6c[dg]Zg c\Zc\j hL dj h] deh

In Abbildung 1 ist das Ergebnis der Diskussion von Anforderungen der mFUND-Projektnehmer zu sehen. Für ein Datenportal, wie es in OPAL untersucht werden soll, sind praktisch nur die "Anforderungen an mCLOUD" (grüne Zettel) relevant, da in OPAL kein unmittelbarer Einfluss auf Datenlieferanten oder Dritte besteht. Die folgenden Anforderungen in Tabelle 8 können direkt übernommen oder abgeleitet werden.

Anforderung	Art	Beschreibung
AW1	F	Semantische Suche. Mit dieser Anforderung wird ein umfangreiche Suchfunktionalität (siehe AF1) über den verknüpften und einheitlichen Daten (AF3) mit hoher Qualität (AF8) ermöglicht.
AW2	F	Räumliche Suche. Die typischerweise räumlich begrenzten Daten aus mCLOUD/MDM und verwandten Portalen müssen für die Suchfunktionalität (AF1) geeignet gefiltert werden können.
AW3	F	Zeitliche Suche / Newsletter. Über das kontinuierliche Crawling (AF5) und die extrahierten und in ein einheitliches Format transformierten Metadaten wird die Einschränkung auf Zeiträume (AW3.1) und ggf. die Information von Nutzern über neue oder aktualisierte Inhalte ("Newsletter", AW3.2) ermöglicht.

I **WiaZ**<sup>°</sup> -: Anforderungen aus dem mFUND-Workshop

Einige "grüne" Anforderungen sind in Tabelle 8 nicht aufgeführt. Die Anforderung "Einheitliche Struktur der Daten" entspricht Anforderung AF3. "Laufzeit der Datenbereitstellung sichern" ermöglicht ein bessere Dokumentation des Datensatzes, liegt aber außerhalb des Einflussbereiches von OPAL. "Vernetzung der Nutzer" und "Wiki mit Forumsfunktion" erfordern umfangreichere Community-Funktionalität, die nicht Ziel von OPAL ist - OPAL wird hier

## D1.1 - Anforderungsanalyse

höchstens eine "Nutzerhistorie der Datenquellen" (AW4) ermöglichen. "Regeln der Datennutzung" wird durch die Anforderung AF7 zur Lizenzintegration abgedeckt.

Weitere Informationen zu den Ergebnissen des Workshops sind auf der Seite des Veranstalters WIK zu finden.<sup>11</sup>

### (# 7Zhi Z] ZcYZ°6c[dgYZg c\ZcVj hYZg: ci l X aj c\ kdc b 8ADJ 9'

In der "Fachkonzeption zum BMVI-Datenportal" (vertraulich) wurden insbesondere in Abschnitt 2 Anforderungen und Vorgaben für mCLOUD zusammengefasst. Interessant ist zunächst die Abgrenzung gegenüber bestehenden Portalen, die erläutert, dass mCLOUD den aktiven Austausch mit dem Nutzer fördert, teilweise auch nicht-offene Daten auflistet und auf bereits existierende Infrastrukturen zurückgreifen soll, also nicht selbst Daten bereitstellt.

Die Anforderungen sind in 5 Gruppen unterteilt. Im Hinblick auf die für OPAL besonders relevanten Metadaten fordert die Fachkonzeption lediglich, dass die Daten über Metadaten beschrieben werden - im "Optimalfall [...] über standardisierte Formate". Weiterhin sollen die Nutzungsbedingungen langfristig der Geodatennutzungsverordnung des Bundes (GeoNutzV) entsprechen. In der weiteren Diskussion wird festgestellt, dass Metadaten je nach Anbieter sehr unterschiedlich vorliegen, z.B. bei Geoinformationen durch ISO-Standards, während Klimadaten "durch die Benamung von Dateien und Verzeichnissen strukturiert" sind, ebenso werden fehlende Lizenzinformationen erkannt. In OPAL-Deliverable D1.2 wird dies detailliert untersucht.

Weiterhin wurde das mCLOUD-Portal in einer Usability-Studie untersucht. Wichtige Erkenntnisse, die als Anforderungen an die Suchfunktionalität in OPAL betrachtet werden sollten, sind:

- Unterstützung unterschiedlicher Schreibweisen der Suchbegriffe (derzeit unterschiedliche Ergebnismengen für "Ladestation", "Ladestationen", "Ladestation\*" durch fehlerhaftes Stemming)

- Facettierte Suche mit Unterstützung für UND- und ODER-Verknüpfungen (derzeit nur UND)

- Umfangreiche Metadatenbeschreibung hinsichtlich Auflösung, Genauigkeit sowie Stand und Aktualität, Verwendung von DCAT-AP/DCAT-GEO

- API

- Kartenviewer

- Veränderungshistorie

### (# : ^ZcZ°6c[dgYZg c\Zc'

Zusätzlich zu den bereits zuvor genannten Anforderungen ergeben sich die folgenden Anforderungen aus der Vorhabensbeschreibung und den in OPAL betrachteten Anwendungsszenarien (insbesondere AP7 und AP8).

Anforderung	Art	Beschreibung
AV1	F	Automatisierte Extraktion von Metadaten aus relevanten Datenquellen

<sup>11</sup>

[http://www.wik.org/fileadmin/Sonstige\\_Dateien/mFund/20171218\\_WIK\\_mFUND\\_Auftaktworkshop\\_ErgebnisseAGn\\_IN.pdf](http://www.wik.org/fileadmin/Sonstige_Dateien/mFund/20171218_WIK_mFUND_Auftaktworkshop_ErgebnisseAGn_IN.pdf)

## D1.1 - Anforderungsanalyse

		anhand von Beispieldaten
AV2	F	Fokussierter Crawler, der relevante Webseiten mit offenen Daten anhand bereitgestellter Seed-Listen findet und untersucht (siehe auch AE1 in Tabelle 4)
AV3	F	Effizientes periodisches Recrawling bereits untersuchter Quellen (siehe auch AF5 in Tabelle 7)
AV4	F	Automatische Berechnung von generischen Qualitätsmetriken (siehe auch AE5 und AF4)
AV5	F	Spezifische Qualitätskriterien berechnet aus impliziten unstrukturierten Metadaten
AV6	F	Konvertierung der Metadaten in verschiedene Formate, ggf. mit unterschiedlichen Vokabularen
AV7	F	Verknüpfung von Datensätzen anhand von Ähnlichkeitsmetriken auf Metadaten (siehe auch AE10)
AV8	F	Möglichkeit der Selektion von Teilmengen eines Datensatzes
AV9	F	Mobile Anwendung zur Suche nach lokal relevanten Datensätzen
AV10	F	Question-Answering-basierter Assistent für soziale Netzwerke
AV11	F	Untersuchung der Datensätze nach Datenportal, Datenanbieter, und weiteren Facetten, die mit logischen UND- und ODER-Operatoren kombiniert werden können
AV12	F	Statistiken und Zeitverlauf sowie Versionierung zu Datensätzen

**IVWaz**: Anforderungen aus den OPAL-Anwendungsszenarien

### ) Q hMb b Zc[Vrhj c\j cY°Gj hWIX

Im vorliegenden Projektbericht sind für das OPAL-Projekt Anforderungen erfasst und analysiert worden. Hierzu wurden zunächst die möglichen Quellen von Anforderungen identifiziert und anschließend systematisch untersucht.

### )#& @dchdaY'ZgZ°6c[dgYZg c\Zc'

Die vorher separat erfassten Anforderungen wurden in der folgenden Tabelle 10 konsolidiert, um einen Überblick über die Anforderungen an OPAL zu erhalten. Die Einzelanforderungen sind in der vierten Spalte referenziert (siehe vorhergehende Tabellen).

Anforderung	Art	Beschreibung	referenziert
AK1	F	Semantische Suche. Mit dieser Anforderung wird ein umfangreiche Suchfunktionalität über den verknüpften und einheitlichen Daten mit hoher Qualität ermöglicht.	AW1, AF1, AF3, AF8



## D1.1 - Anforderungsanalyse

AK2	F	Räumliche Suche. Die typischerweise räumlich begrenzten Daten aus mCLOUD/MDM und verwandten Portalen müssen für die Suchfunktionalität (AF1) geeignet gefiltert werden können.	AW2, AF1
AK3	F	Zeitliche Suche. Über das kontinuierliche Crawling und die extrahierten und in ein einheitliches Format transformierten Metadaten wird die Einschränkung auf Zeiträume ermöglicht.	AW3, AF5, AV3
AK4	F	Monitoring und Benachrichtigung. Das kontinuierliche Crawling ermöglicht die Überwachung der Verfügbarkeit der Daten sowie die Information von Nutzern über Zeitverlauf und Aktualisierungen.	AW3, AE13, AF5, AV3, AV12
AK5	F	Entwicklung eines komponentenbasierten Metadatenportals.	AF1
AK6	F	Programmatischer Zugriff über standardisierte API. Wenn möglich sollen mehrere Zugriffsmethoden und Datenformate bereitgestellt werden, um den unterschiedlichen Nutzeranforderungen zu genügen.	AE11, AF2,
AK7	F	Konvertierung der Metadaten in Linked Data und unterschiedliche Formate und Vokabulare	AE3, AF3, AV6
AK8	F	Analyse der (Meta-)Datenqualität. Die Qualität der Daten wird mit geeigneten generischen und spezifischen Metriken untersucht.	AE5, AE12, AV4, AV5, AF4
AK9	F	Automatisierte Verknüpfung von Datensätzen.	AE10, AF6, AV7
AK10	F	Prüfung und Integration von Lizenzinformationen. Motivation ist die hohe Bedeutung korrekter Lizenzinformationen, als auch die Notwendigkeit, resultierende Lizenzbedingungen aus der Integration von mehreren Datensätzen abzuleiten.	AF7
AK11	F	Fokussierter Crawler, der relevante Webseiten mit offenen Daten anhand bereitgestellter Seed-Listen findet und untersucht	AE1, AV2
AK12	F	Automatisierte Extraktion von Metadaten aus relevanten Datenquellen anhand von Beispieldaten	AV1
AK13	F	Möglichkeit der Selektion von Teilmengen eines Datensatzes für ausgewählte Datenquellen	AV8
AK14	F	Mobile Anwendung zur Suche nach lokal relevanten Datensätzen	AV9
AK15	F	Question-Answering-basierter Assistent für soziale	AV10

## D1.1 - Anforderungsanalyse

		Netzwerke	
AK16	F	Untersuchung der Datensätze nach Datenportal, Datenanbieter, und weiteren Facetten, die mit logischen UND- und ODER-Operatoren kombiniert werden können	AV11
AK17	F	Persistente versionierte Speicherung von Metadaten	AE2
AK18	F	Anzeige von existierenden und neuen Daten im Repository, einschließlich verfügbarer Metadaten.	AE4
AK19	F	Empfehlungen von relevanten Datensätzen basierend auf semantischer Dokumentenanalyse, Suchanfrage, gewählten Datensätze und persönlichem Profil.	AE7
AK20	F	Kommentierung bezüglich Qualitäts- und inhaltsbezogener Kriterien.	AE8
AK21	F	Möglichkeit zur Bewertung der Datensätze anhand von nutzerzentrierten Qualitätskriterien.	AE9

### 1.1.2 Konsolidierte Anforderungen an OPAL

#### 1.1.2.1 Einführung

Die beschriebene Erfassung und Analyse der Anforderungen bildet die Grundlage für die folgenden Projektberichte zu initialer Datenanalyse und der Entwicklung der OPAL-Architektur. Weiterhin stellt sie eine Grundlage für die Spezifikation und Entwicklung der Komponenten in den jeweiligen Arbeitspaketen 2 bis 8 dar.