

]init[



WIG 2

# Datennutzung in der Post-COVID- Forschung

Bessere Auffindbarkeit von Forschungsprojekten durch  
standardisierte Metadaten



# Herzlich willkommen zur Auftaktveranstaltung der Veranstaltungsreihe zum Thema "Datennutzung in der Post-COVID-Forschung"!

## Allgemeine Hinweise

### Aufzeichnung & Veröffentlichung

Einige Teile dieses Workshops werden aufgezeichnet und im Internet ([z. B. auf YouTube](#)) veröffentlicht. Mit Ihrer Teilnahme stimmen Sie der Aufzeichnung und Veröffentlichung zu.

### Materialien

Folien und Aufzeichnung werden Ihnen im Nachgang per Mail und auf unserer [Webseite](#) zur Verfügung gestellt.

# Warmup



Warum sind Sie heute beim Workshop dabei?

16 Antworten

dcat subprofile  
munbekannte datenquellen  
vernetzung  
weiterbildung  
analyseoptionen  
health dcat-ap  
interesse an metadaten  
wissensaustausch  
interesse  
standardisierung  
keynote halten  
neugier  
interoperabilität  
kooperationspartner finde  
versorgungsdaten nutzen  
forschung

# Agenda



## Teil 1 des Workshops

- Begrüßung (mit Aufzeichnung)
- Keynotes (mit Aufzeichnung)
- Vorstellung des Metadatenmodells (mit Aufzeichnung)
- Q&A-Runde (mit Aufzeichnung)

## Teil 2 des Workshops

- Interaktive Diskussion in Gruppen (ohne Aufzeichnung)
- Vorstellung der Ergebnisse der Diskussion in Gruppen (mit Aufzeichnung)
- Ausblick auf nächste Schritte & Verabschiedung (mit Aufzeichnung)

# Keynote 1



**Prof. Dr. Wilfried Bernhardt**

Wissenschaftlicher Beirat der AOK Nordost

# Datennutzung im Gesundheitssektor

## Keynote 2



**Sebastian Sklarß**

Experte für EU-Interoperabilität

# Vorstellung des Vorhabens im Rahmen der Post-COVID-Challenge

# Ausgangslage

## Forschung mit Gesundheitsdaten



Problematik des Datenschutzes



Nur schwierig auffindbar



Unterschiedliche Formate



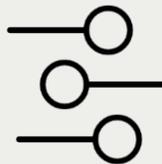
Verknüpfungen aufwändig

# Ausgangslage

## Forschung zu Post-COVID



Internationales Phänomen



Diffuses Krankheitsbild



Schnelllebige Forschung

# Ausgangslage

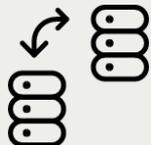
## Nationale Vorhaben



Gesundheitsdatennutzungsgesetz (GDNG)



Digitalisierungsstrategie des BMG



Datenhaltungsinfrastrukturen

# Dateninstitut

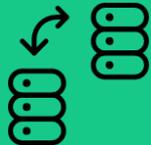
## Nationale Vorhaben



Gesundheitsdatennutzungsgesetz (GDNG)



Digitalisierungsstrategie des BMG



Datenhaltungsinfrastrukturen

# Vision

## Das Dateninstitut



Dateninstitut



Metadatenkatalog



Werkzeugkoffer



Vernetzungsplattform



Suche



Software-  
Bibliothek



Validator



Eingabemaske

# Vision

## Das Dateninstitut



# Vision

## Das Dateninstitut



Dateninstitut



Metadatenkatalog



Werkzeugkoffer



Vernetzungsplattform



Suche



Software-  
Bibliothek

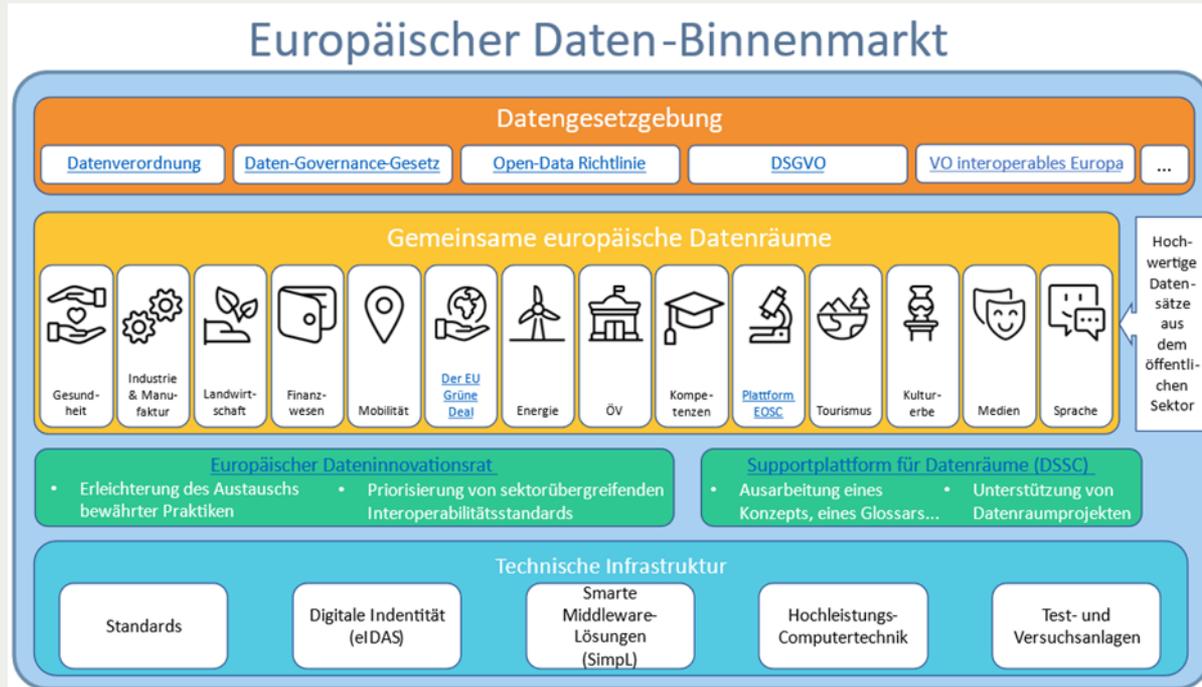


Validator

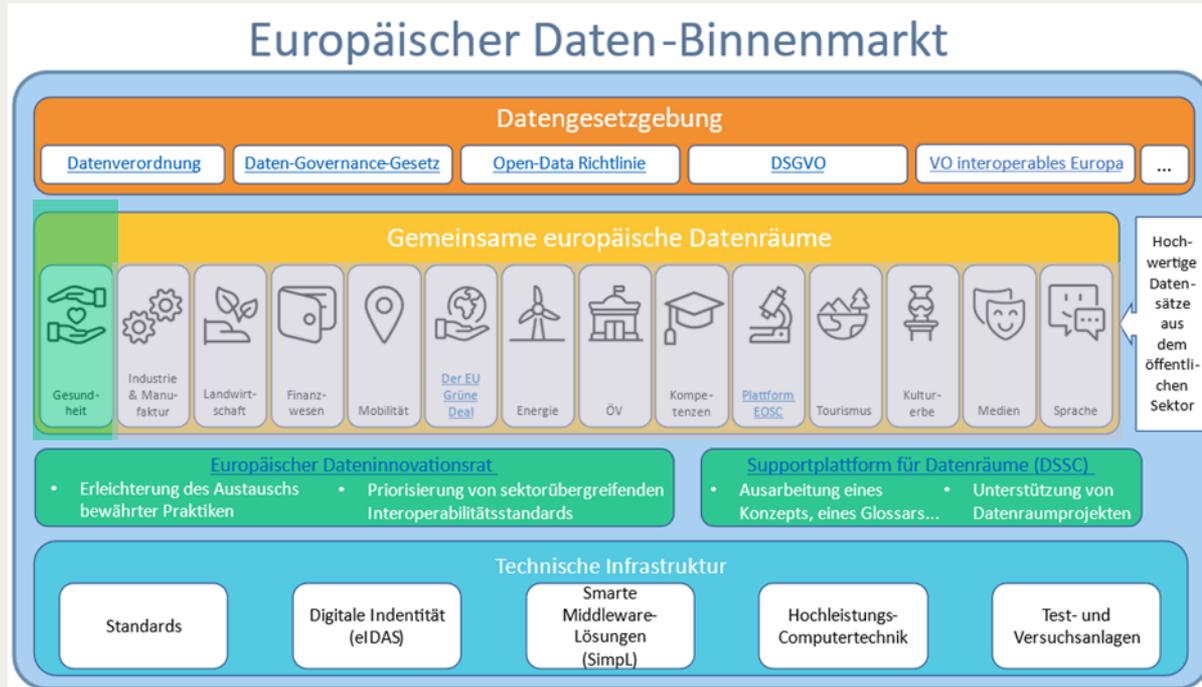


Eingabemaske

# Europäischer Daten-Binnenmarkt



# Europäischer Daten-Binnenmarkt



HealthDCAT-AP

aus  
HealthData@EU

# Post-COVID-Challenge – der Lösungsansatz

## Nutzung eines EU-Kernvokabulars



**DCAT** = W3C-Vokabular für die **Interoperabilität** zwischen Datenkatalogen

**DCAT**  
W3C



**DCAT-AP** = Europäisches Anwendungsprofil für **EU**-Datenportale



**DCAT-AP.de** = deutsches Metadatenmodell für offene Verwaltungsdaten

**StatDCAT-AP** = europäisches Metadatenmodell für Statistische Datensätze



**HealthDCAT-AP** aus HealthData@EU



# In a nutshell

## Der Ansatz von ]init[ und WIG2

Ein Metadatenmodell mit:

#DatenschutzByDesign

#EU-Interoperabilität

#HealthDCAT-AP

#Wissensgraph

# Vorstellung des Metadatenmodells



**Daniel Brenn**

Modellierungsexperte

# Vorstellung des Metadatenmodells HealthDCAT-AP.de

# Umgang mit Daten

## FAIR-Prinzipien



FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE



# Warum Metadaten?

## Erfüllung der FAIR-Prinzipien



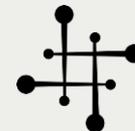
Auffindbarkeit durch Maschinenlesbarkeit



Keine Zugangsbeschränkungen



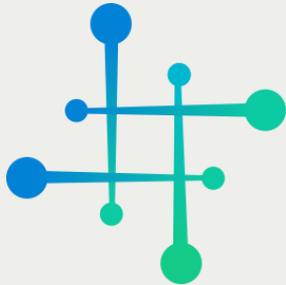
Interoperabilität durch Standardisierung



Nachnutzbarkeit durch offenes Format

# HealthDCAT-AP.de

## Anpassung und Erweiterung von HealthDCAT-AP



HealthDCAT-AP



Ausprägung für **Gesundheitsdaten** von **DCAT-AP**



# HealthDCAT-AP.de

## Anpassung und Erweiterung von HealthDCAT-AP



DCAT-AP.de



Deutsche Ausprägung von **DCAT-AP**



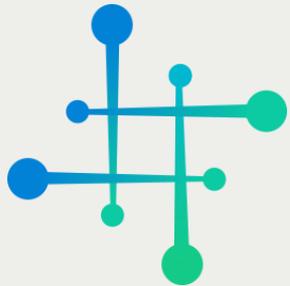
Croissant



Erweiterung von **schema.org** für  
die **KI-optimierte Darstellung** von  
**Metadaten** und **Datenstrukturen**

# HealthDCAT-AP.de

## Anpassung und Erweiterung von HealthDCAT-AP



HealthDCAT-AP



DCAT-AP.de



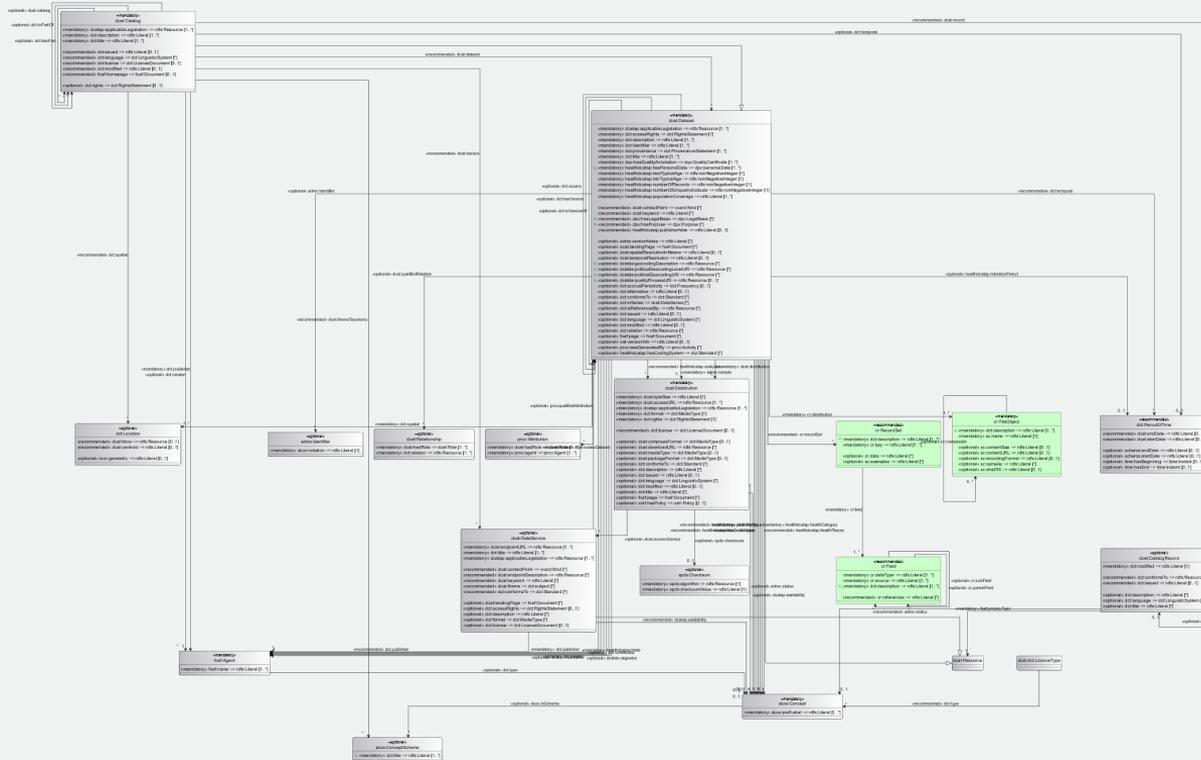
Croissant



HealthDCAT-AP.de

# HealthDCAT-AP.de

## Blick auf das gesamte Datenmodell



# HealthDCAT-AP.de

## Blick auf die wichtigsten Klassen



 dcat:Dataset

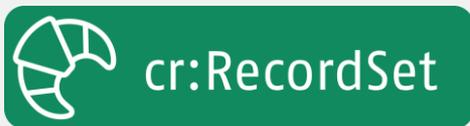
Ein spezifischer Datensatz

 dcat:Distribution

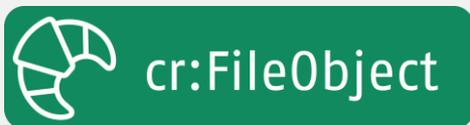
Verkörperung des Datensatzes  
in einem spezifischen Format

# HealthDCAT-AP.de

## Blick auf die wichtigsten Klassen



Sammlung homogener Datenobjekte  
in einem Datensatz

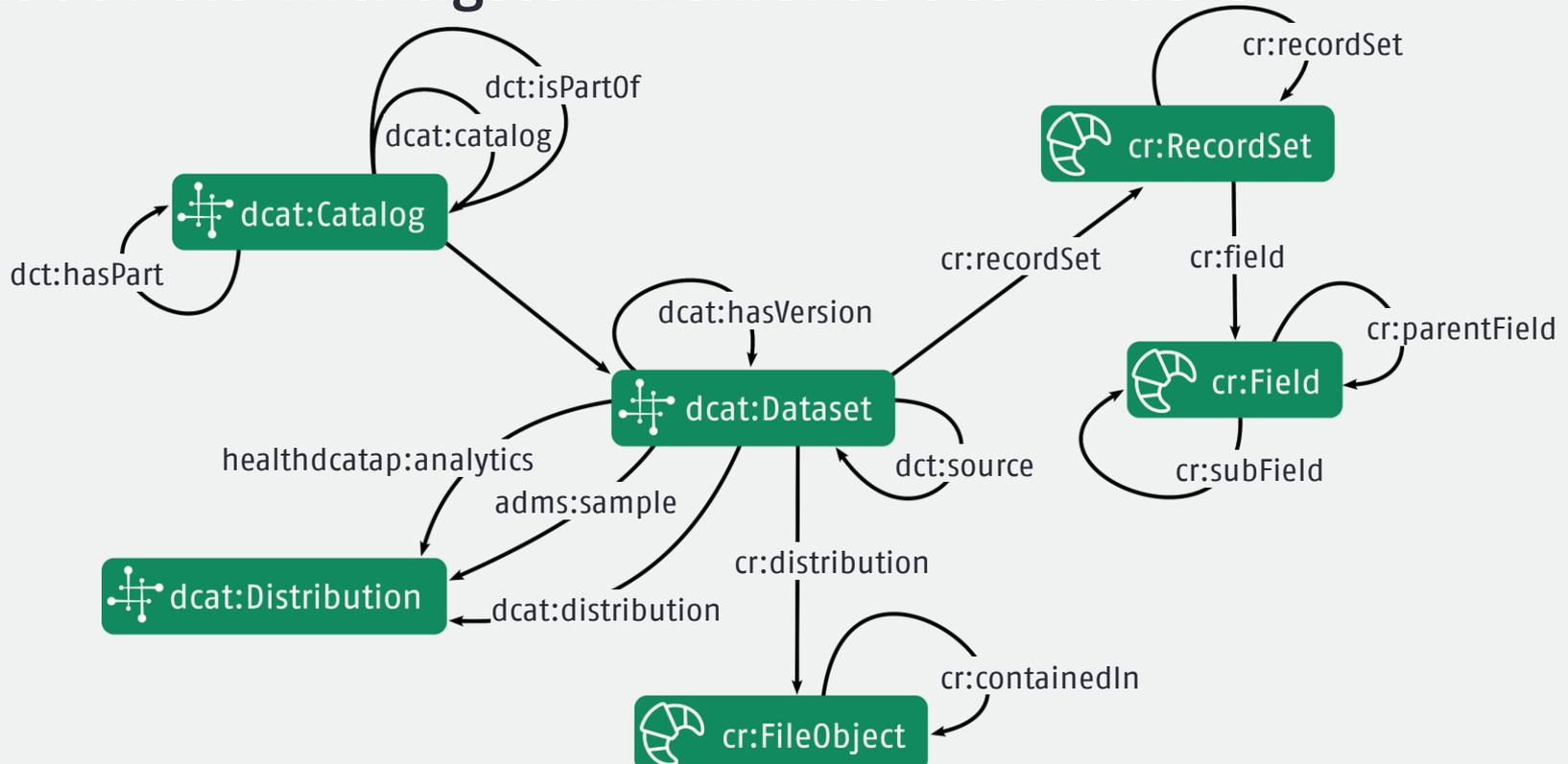


Datei in einem Datensatz

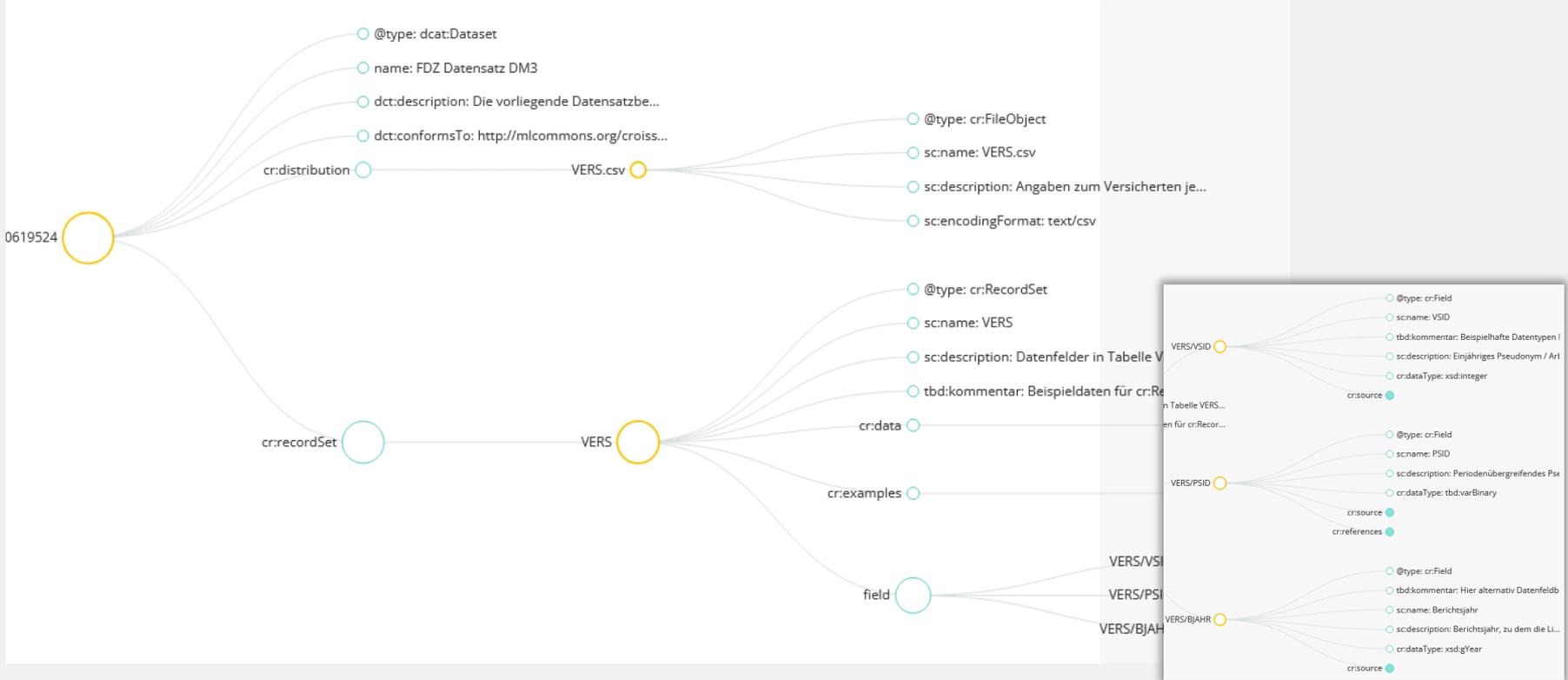


Beschreibung der Datensätze

## Blick auf die wichtigsten Elemente des Modells



# HealthDCAT-AP.de Datensatz



# HealthDCAT-AP.de

## Mehrwert Croissant



Auffindbarkeit durch Maschinenlesbarkeit



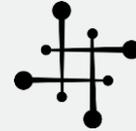
Detaillierte Informationen zur Struktur von Datensätzen

Angabe von konkret genutzten Wertelisten und Normvokabularen vorab erkennbar





Erweiterung DPV ermöglicht Blick auf wichtige  
Datenschutzfragen



Detaillierte Metadaten zeigen vor dem  
eigentlichen Zugriff auf den Datensatz, ob die  
Daten für die eigene Forschung geeignet sind

DCAT-Vokabular ermöglicht  
Sekundärnutzung aus anderen Bereichen:  
Umwelt, Geodaten, usw.



# Datenmodell und Wissensgraph

## Wie können Daten beigetragen werden?



| Klasse     | Erforderlichkeit | Eigenschaft                 | Kardinal | Wertebereich            | Hinweis  | Wert                         | Anmerkung |
|------------|------------------|-----------------------------|----------|-------------------------|--|------------------------------|-----------|
| dc:Dataset | mandatory        | adms:sample                 | [1..*]   | dc:Distribution         | Link zu einem Beispiel(datensatz)  |                              |           |
| dc:Dataset | mandatory        | cr:distribution             | [1..*]   | cr:FileObject           | Namen der zum Datensatz gehörigen FileObjects.<br>Diese Eigenschaft verknüpft den Datensatz mit einer verfügbaren Distribution (Quelle: dc:ap.de). Eine Distribution ist eine konkrete Veröffentlichung des Datensatzes.   | [wird automatisch generiert] |           |
| dc:Dataset | mandatory        | dc:theme                    | [1..*]   | skos:Concept            | <a href="#">Eine Hauptkategorie der zugrundeliegenden Daten. Es sollte mindestens das Label HEALTH verwendet werden. Bitte wählen Sie ein oder mehrere Label aus dem Tabellenblatt Wl_Data theme.</a><br>Die Rechtsvorschriften, die die Erstellung oder Verwaltung des Datensatzes vorschreiben(vgl. dc:ap.de).<br>Bitte verweisen Sie für die URL auf die einschlägige Seite auf <a href="http://www.gesetze-im-internet.de">www.gesetze-im-internet.de</a> .<br>Zusätzlich zur URL kann ein Label und ein Kommentar angegeben werden. |                              |           |
| dc:Dataset | mandatory        | dc:ap:applicableLegislation | [1..*]   | rdfs:Resource           | <a href="#">Diese Eigenschaft verweist auf Informationen, die darlegen, ob der Datenservice öffentlich zugänglich ist. Zugriffsbeschränkungen existieren oder er nicht-öffentlich ist. Bitte wählen Sie einen Code aus der Werteliste Wl_Access right.</a>   |                              |           |
| dc:Dataset | mandatory        | dct:accessRights            | [1]      | dct:RightsStatement     | Freitextbeschreibung des Datensatzes   |                              |           |
| dc:Dataset | mandatory        | dct:description             | [1..*]   | rdfs:Literal            | Ein oder mehrere persistente URI zum Datensatz.  |                              |           |
| dc:Dataset | mandatory        | dct:identifier              | [1..*]   | rdfs:Literal            | Diese Eigenschaft umfasst eine Angabe zur Entwicklungsgeschichte des Datensatzes, sowie Methoden, Tools und Konventionen, welche bei der Datenerhebung berücksichtigt wurden. Insbesondere ist relevant, in wessen Besitz oder Obhut die Ressource sich bislang befunden hat, soweit die Wechsel signifikanten Einfluss auf die Authentizität, Integrität und Interpretierbarkeit dieser Ressource hat (vgl. dc:ap.de, healthdcat-ap).   |                              |           |
| dc:Dataset | mandatory        | dct:provenance              | [1..*]   | dct:ProvenanceStatement | Link zu Onlinepräsenz der Organisation, welche den Datensatz zur Verfügung stellt.<br><a href="#">Räumlicher Bezug des Datensatzes. Bitte Label aus dem Tabellenblatt WK_Countries...wählen (Deutschland = DEU)</a>  |                              |           |
| dc:Dataset | mandatory        | dct:publisher               | [1]      | foaf:Agent              | Ein menschenlesbarer Titel des Datensatzes<br><a href="#">Bezieht sich auf den Typ des Datensatzes. Bitte wählen Sie ein Label aus der Werteliste Wl_Dataset type.</a>   |                              |           |
| dc:Dataset | mandatory        | dct:spatial                 | [1..*]   | dct:Location            | Eine Anmerkung, die eine Ressource (insbesondere einen Datensatz oder eine Verteilung) mit einer   |                              |           |
| dc:Dataset | mandatory        | dct:title                   | [1..*]   | rdfs:Literal            |  |                              |           |
| dc:Dataset | mandatory        | dct:type                    | [1]      | skos:Concept            |  |                              |           |



### HealthDCAT-AP.de

**Vielen Dank** | Vielen Dank für Ihre Unterstützung des Vorhabens HealthDCAT-AP.de, ihre Metadaten ermöglichen uns, die Mehrwerte des Wissensgraphen und Funktionsweise des Metadatenstandards HealthDCAT-AP.de zu

**Hinweis zur Bearbeitung** | Im Rahmen der Demonstration erfassen wir Metadaten für drei Objektklassen, Dataset, DataService und FileObject. Für jede Klasse haben wir ein Tabellenblatt angelegt und auf die im Datenmodellentwurf mandatorischen Angaben vorgefildert.  
Wir wissen, dass einige Angaben voraussetzungs- und/oder initierend sein können. In diesem Fall bitten wir Sie, uns ohne Zögern telefonisch oder per Mail zu kontaktieren.

**Das Vorhaben** | Die Init AG für digitale Kommunikation und des WIG2 Institut für Gesundheitswissenschaften, nehmen an der Endrunde der Post-COVID-Challenge zur Gründung des Dateninstituts des Bundes teil. Gemeinsam nutzen wir unsere langjährigen Erfahrungen mit interoperabler Datenmodellierung (init) und gesundheitsökonomischer Forschung (WIG2) für die Entwicklung des deutschen Profils des europäischen Metadatenstandards für den europäischen Gesundheitsraum, HealthDCAT-AP.de.  
Die Challenge markiert den Startschuss für das Dateninstitut, das u.a. zusammen mit dem Bundesforschungsministerium zu einer besseren Nutzung von Daten beitragen soll. Ziel der Challenge ist, sowohl die

**Weitere Informationen** | [healthdcat-ap.de@healthdcat-ap.de](mailto:healthdcat-ap.de@healthdcat-ap.de)

**Rückfragen** | **Jonas Pauly**  
jonas.pauly@init.de  
+49 331 7426 7314

**Robin Müller**  
robin.mueller@init.de

**Kontakt** | [pf.healthdcat-ap@init.de](mailto:pf.healthdcat-ap@init.de)



# Datenmodell und Wissensgraph

## Wie können Daten beigetragen werden? Ausblick

The screenshot displays the 'Edit Dataset' page in the HealthDCAT-AP system. It features a multi-step progress indicator at the top, with steps 1 through 5. Step 1, 'Essential Properties', is currently active. Below the progress bar, there are three main sections for editing dataset metadata: 'Title of the Dataset', 'Dataset ID\*', and 'Description of the Dataset'. Each section has a text input field and a small icon. To the right of these sections are buttons for 'Publish Dataset', 'Save as Draft', and 'Next Step'. Below the metadata sections is a 'Distribution Overview' section, which lists existing distributions with their URIs and icons for adding or removing them. The interface is clean and modern, with a blue header and footer.

# Lösungsvorschlag

## Was machen wir in den einzelnen Stufen?

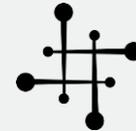


Stufe 1: Entwicklung der Konzepte und des Datenmodells bis Version 0.3



Stufe 2: Weiterentwicklung des Datenmodells, Entwicklung PoC und Durchführung der ersten Workshops

Stufe 3: Etablierung der Workshopreihen, Entwicklung MVP und Schritte zur Verstetigung des Datenstandards





# Austausch zum Metadatenmodell HealthDCAT-AP.de

# Vorstellung der Ergebnisse



**Dr. Felix Pickhardt**

Gesundheitsdatenexperte

# Vorstellung der Ergebnisse der Diskussion in Gruppen



# Vielen Dank für Ihre Teilnahme!

## Folgetermine

Thema Standardisierung: 1. Workshop am 28.11.2024

Thema Datennutzung: 2. Workshop am 13.02.2025

## Folien und Aufzeichnung

Stellen wir Ihnen per Mail und auf unserer [Webseite](#) zur Verfügung.

## Kontakt

Bei Fragen und/oder Interesse zur Mitarbeit melden Sie sich gern per Mail:

[pf-healthdcat-ap@init.de](mailto:pf-healthdcat-ap@init.de)