

# Enabling Open Innovation through Open Data with the JValue Open Data Service

**Prof. Dr. Dirk Riehle**

**Friedrich-Alexander-Universität Erlangen-Nürnberg**

**Universiti Sains Malaysia – 2019-09-04**

# Open Innovation

- Open innovation is
  - “a distributed innovation process based on
  - purposively managed knowledge flows
  - across organizational boundaries” [1]
- Organizations can be
  - The state
  - Companies
  - Non-profits

[1] Chesbrough, H., & Bogers, M. 2014. [Explicating open innovation: Clarifying an emerging paradigm for understanding innovation](#). In H. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *New Frontiers in Open Innovation*: 3-28. Oxford: Oxford University Press. Page 17.

# Open Data

- Open data
  - is data that
  - “can be freely used, modified, and shared
  - by anyone
  - for any purpose” [1]
- Data is marked as open by being provided under a conformant license [2]

[1] See <https://opendefinition.org>

[2] See <https://opendefinition.org/licenses/>

# Publishers of Open Data

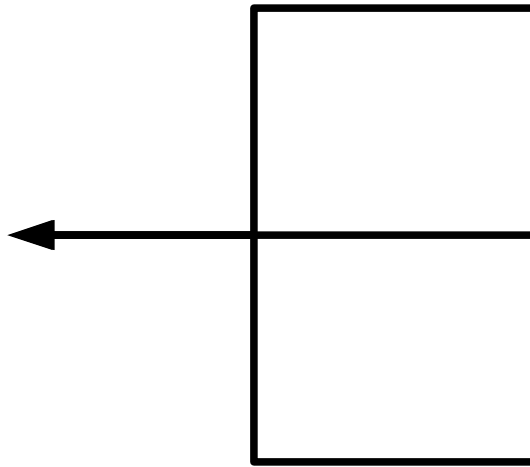
- Governmental agencies (often forced by law)
  - For example, destatis (German statistical data)
    - Example application: Visualization of crime statistics on a map
- Companies interested in fostering open innovation
  - For example, Deutsche Bahn (main German railway operator)
    - Example application: Visualization of data network strength along railway tracks
- Non-profit organizations trying to fulfill their mission
  - For example, Coding da Vinci (German museums and libraries)
    - Example application: Experience enrichment by combining museum data with Wikipedia
- Often provided only in spirit, because a license is missing

# Open Data as an Enabler of Open Innovation

## Innovative App



**PegelAlarm**



## Open Data Sources



**OpenStreetMap**



**PEGELONLINE**



**Deutscher Wetterdienst**

# Some Challenges to Using Open Data

- Technical challenges
  - Data access
    - Plethora of protocols and formats
    - Poor QoS of data publishers
  - Data quality
    - Faulty data (invalid, incomplete)
    - No schema information
  - Data integration
    - No agreed-upon data types
    - No same-entity recognition
- Social challenges
  - Legal issues
    - Unclear data lineage
    - Unclear data license
  - Lack of cooperation
    - Unwillingness to provide
    - Overreaching ambition

# Example Technical Challenges (DWD)

Index of /

<a href="#">../</a>	
<a href="#">climate_environment/</a>	06-Feb-2019 12:24
<a href="#">test/</a>	30-Jan-2019 14:11
<a href="#">weather/</a>	26-Jul-2019 06:51
<a href="#">README.txt</a>	25-Jul-2017 07:00

Index of /climate\_environment/

<a href="#">../</a>	
<a href="#">CDC/</a>	13-Jul-2019 12:38
<a href="#">GPCC/</a>	14-Feb-2019 07:25
<a href="#">REA/</a>	27-May-2019 13:30
<a href="#">health/</a>	11-Dec-2017 20:46

Index of /climate\_environment/CDC/

<a href="#">../</a>	
<a href="#">derived_germany/</a>	21-May-2019 13:05
<a href="#">grids_europe/</a>	27-Dec-2018 13:37
<a href="#">grids_germany/</a>	23-Nov-2018 10:22
<a href="#">help/</a>	26-Jul-2019 06:05
<a href="#">observations_germany/</a>	24-Oct-2018 08:40
<a href="#">observations_global/</a>	06-Feb-2019 12:55
<a href="#">regional_averages_DE/</a>	19-Nov-2018 08:29
<a href="#">Announce_log_CDC_ftp.txt</a>	29-Jan-2019 10:49
<a href="#">Change_log_CDC_ftp.txt</a>	29-Jan-2019 10:49
<a href="#">Error_log_CDC_ftp.txt</a>	29-Jan-2019 10:49
<a href="#">LiesMich Intro CDC-FTP.pdf</a>	13-Jul-2019 12:38

Index of /climate\_environment/CDC/observati

<a href="#">../</a>	
<a href="#">historical/</a>	26-Jul-2019 06:42
<a href="#">meta_data/</a>	25-Jul-2019 07:09
<a href="#">now/</a>	26-Jul-2019 06:42
<a href="#">recent/</a>	26-Jul-2019 03:31

Index of /climate\_envirc

<a href="#">10minutenwerte TU 13674 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13675 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13696 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13700 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13710 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13711 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13713 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13777 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 13965 now.zip</a>	20-Jul-2019 05:41
<a href="#">10minutenwerte TU 15000 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 15207 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 15444 now.zip</a>	26-Jul-2019 06:42
<a href="#">10minutenwerte TU 15555 now.zip</a>	26-Jul-2019 06:42
<a href="#">BESCHREIBUNG obsgermany climate 10min tu now de.&gt;</a>	13-Jul-2019 12:38
<a href="#">DESCRIPTION obsgermany climate 10min tu now en.pdf</a>	13-Jul-2019 12:38
<a href="#">zehn now tu Beschreibung Stationen.txt</a>	26-Jul-2019 06:42

produkt\_zehn\_now\_tu\_20190726\_20190727\_15555.txt

	STATIONS_ID	MESS_DATUM	QN	PP	TT	TM5	RF	TD	eor
1									
2	15555	201907260000	2	-999	23.8	16.9	51.2	13.	
3	15555	201907260010	2	-999	23.8	16.8	51.3	13.	
4	15555	201907260020	2	-999	23.3	16.9	53.3	13.	
5	15555	201907260030	2	-999	23.1	16.8	54.7	13.	
6	15555	201907260040	2	-999	23.0	17.0	55.7	13.	
7	15555	201907260050	2	-999	22.7	17.0	57.3	13.	
8	15555	201907260100	2	-999	22.1	17.0	60.2	14.	
9	15555	201907260110	2	-999	21.7	17.8	62.6	14.	
10	15555	201907260120	2	-999	20.9	17.8	66.7	14.	
11	15555	201907260130	2	-999	20.9	17.3	67.9	14.	
12	15555	201907260140	2	-999	21.0	17.2	68.3	14.	
13	15555	201907260150	2	-999	21.9	18.4	64.6	14.	
14	15555	201907260200	2	-999	22.1	19.3	64.0	15.	
15	15555	201907260210	2	-999	21.7	19.5	64.8	14.	
16	15555	201907260220	2	-999	20.8	19.1	68.6	14.	
17	15555	201907260230	2	-999	20.7	19.0	68.7	14.	
18	15555	201907260240	2	-999	20.8	19.1	68.7	14.	
19	15555	201907260250	2	-999	20.8	18.8	69.7	15.	
20	15555	201907260300	2	-999	21.8	19.5	64.8	14.	

# Consequence of Technical Challenges to Using Open Data

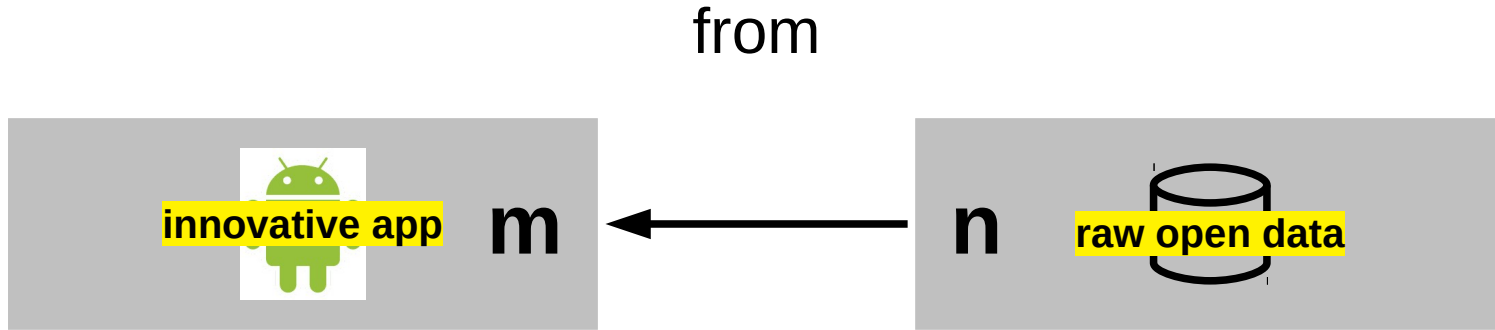
- Data access
  - Plethora of protocols and formats
  - Poor QoS of data publishers
- Data quality
  - Faulty data (invalid, incomplete)
  - No schema information
- Data integration
  - No agreed-upon data types
  - No same-entity recognition



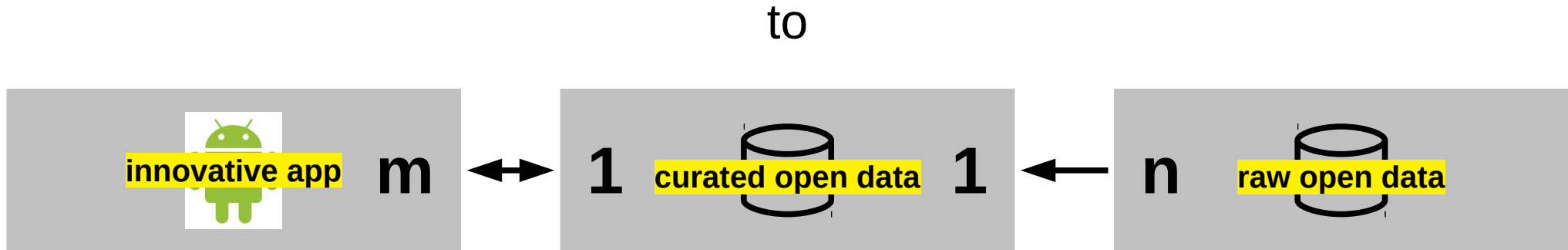
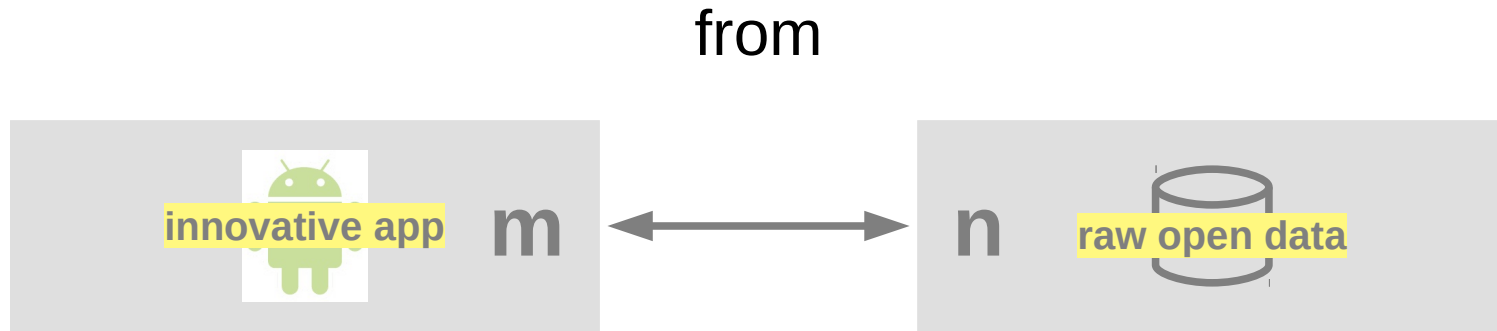
- App development teams spend up to 80% of project time on data gathering, cleaning, preparation
- **Only 20% of time remains for the actual innovation**, turning the open data into useful information



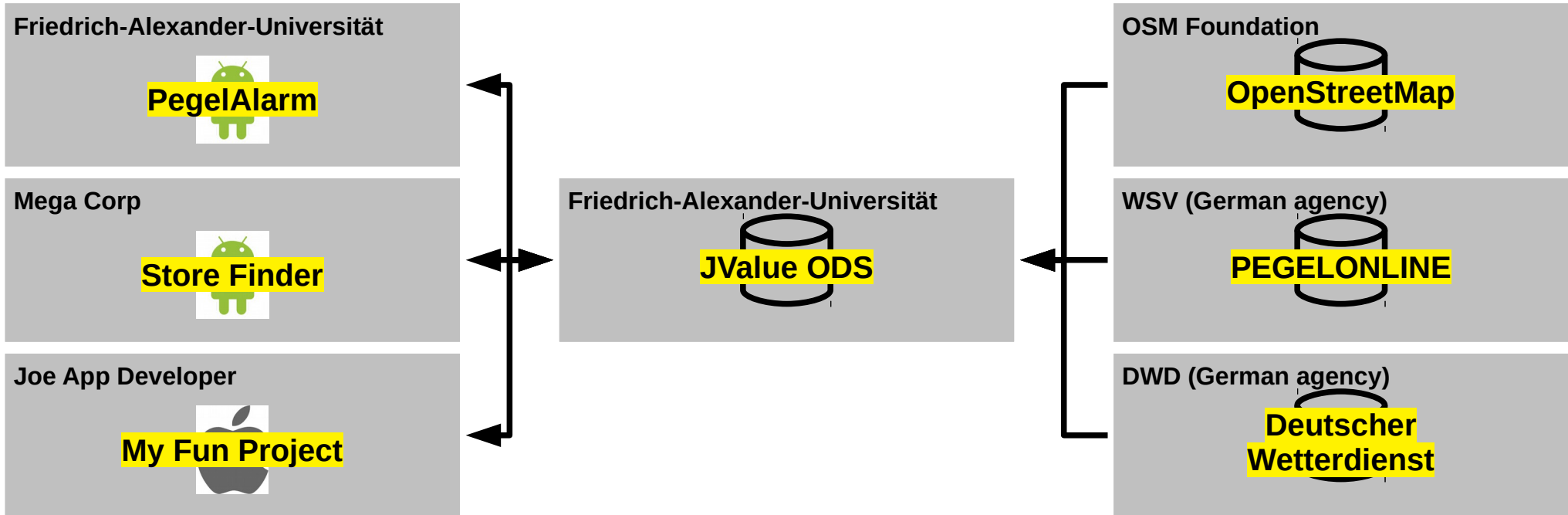
# A Smarter Solution 1 / 2



# A Smarter Solution 2 / 2



# The JValue Open Data Service (ODS)

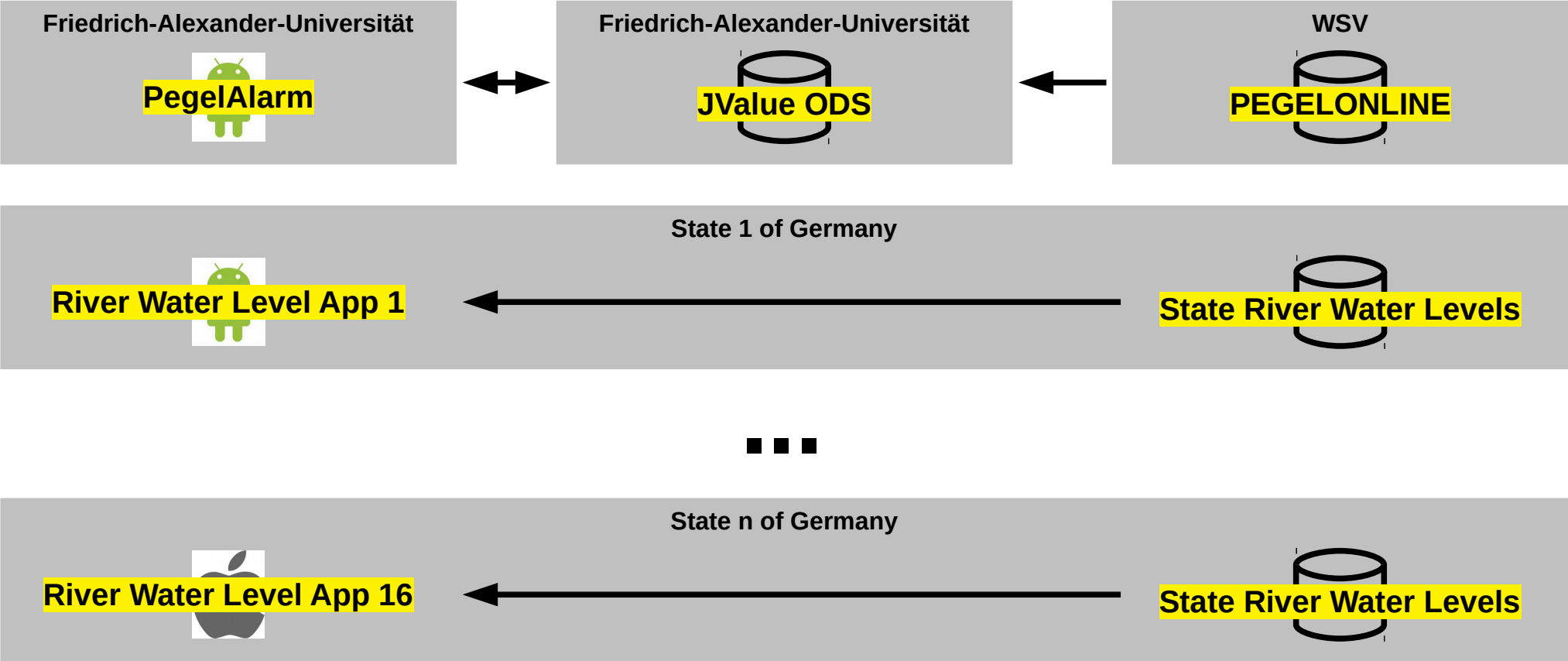


**To make consumption of open data  
easy, reliable, and safe**


# The Project: Both Software and Service

- Software (since 2014!)
  - Open source software (AGPLv3)
  - Available at <https://github.com/jvalue>
- Service
  - A FAU / Prof OSS operated service
  - Example app at <http://pegelalarm.de/>

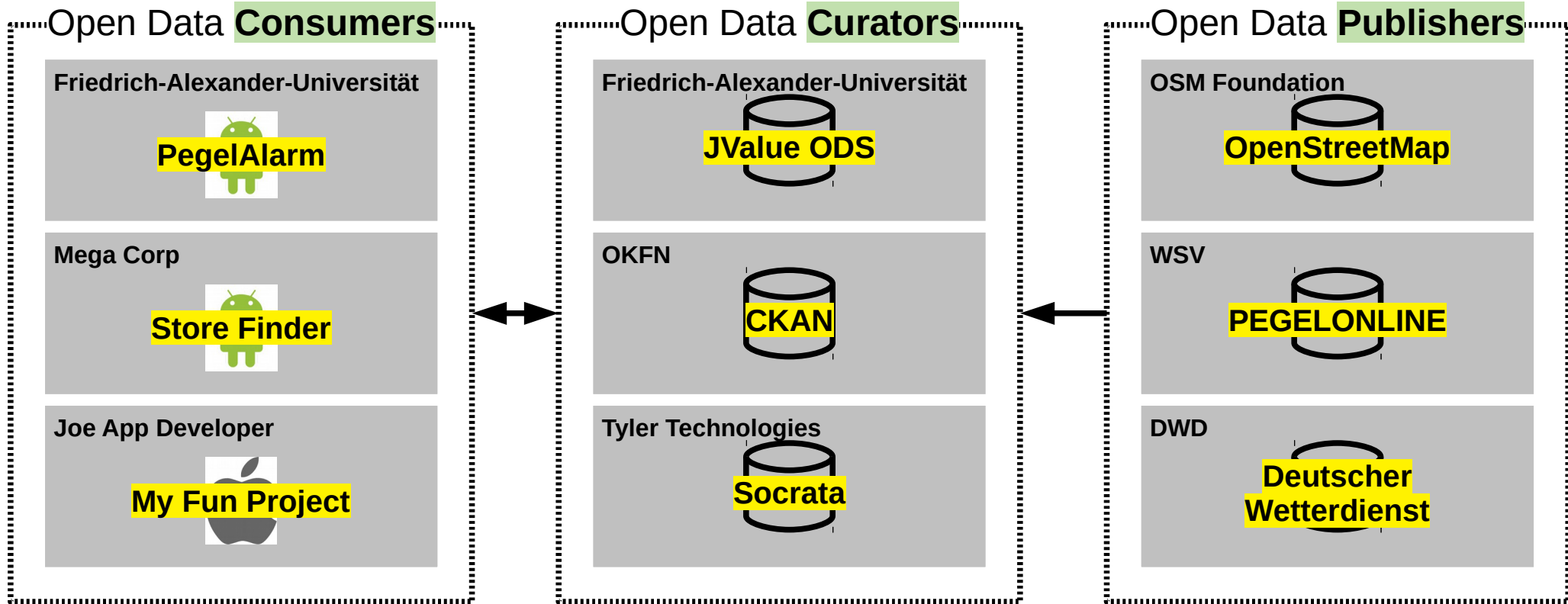
# Example Social Challenges (Silo Applications)



# Consequence of Social Challenges to Using Open Data

- Legal issues
    - Unclear data lineage
    - Unclear data license
  - Lack of cooperation
    - Unwillingness to provide
    - Overreaching ambition
- 
- Data remains hidden (not-open) or is not machine-processable easily
  - Leads to redundant software development efforts (silo applications)
  - Hinders / prevents open innovation

# A Better Open Data Ecosystem





# Components and Their Roles

- Publisher (component)
  - Publishes original open data
  - Plays supplier role to a client
- Curator (component)
  - Plays client role to a supplier
  - Curates from open data publisher
  - Can be chained for value-adding
  - Plays supplier role to a client
- Consumer (component)
  - Consumes data for a purpose
  - Plays client role to a supplier

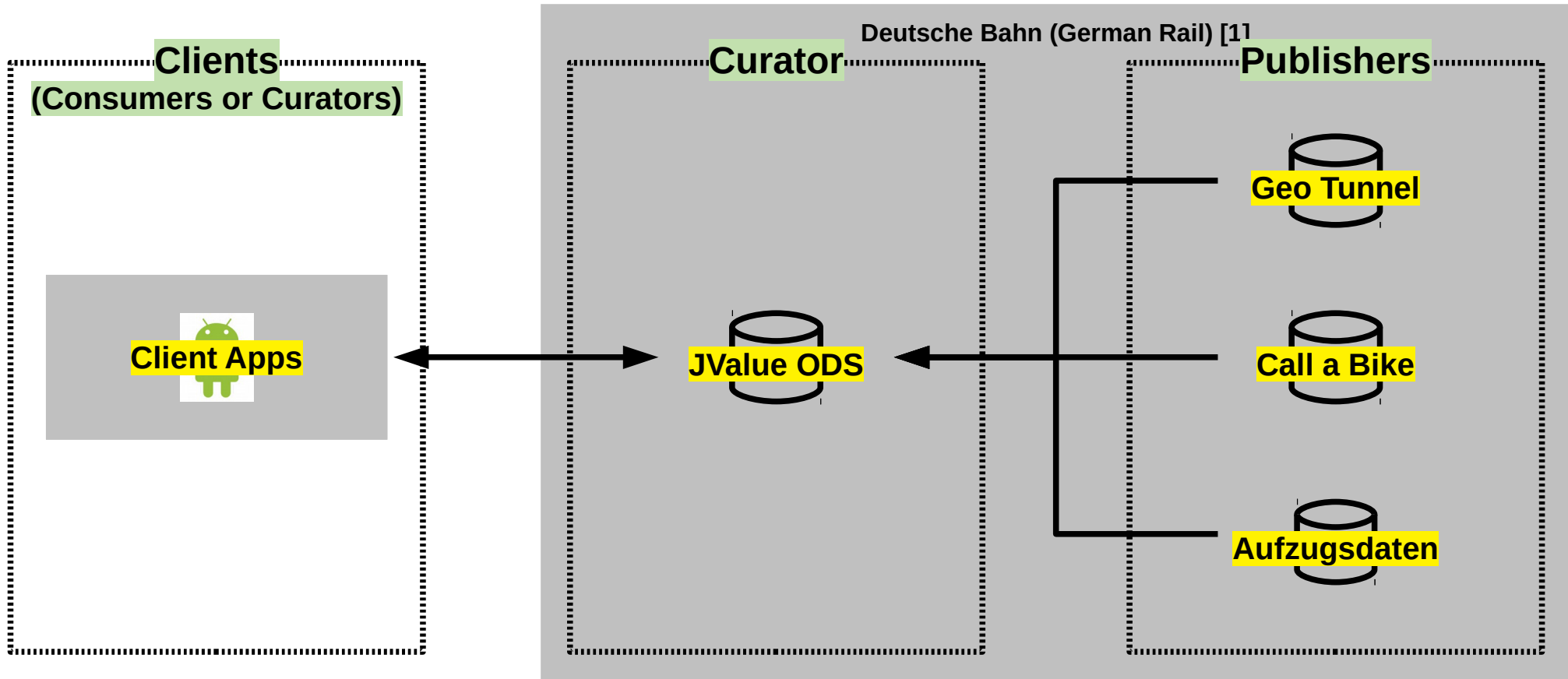
# Relationship Between Software Package Cardinality

**Consumers >> Publishers >> Curators**

# Benefits of Ecosystem Structure

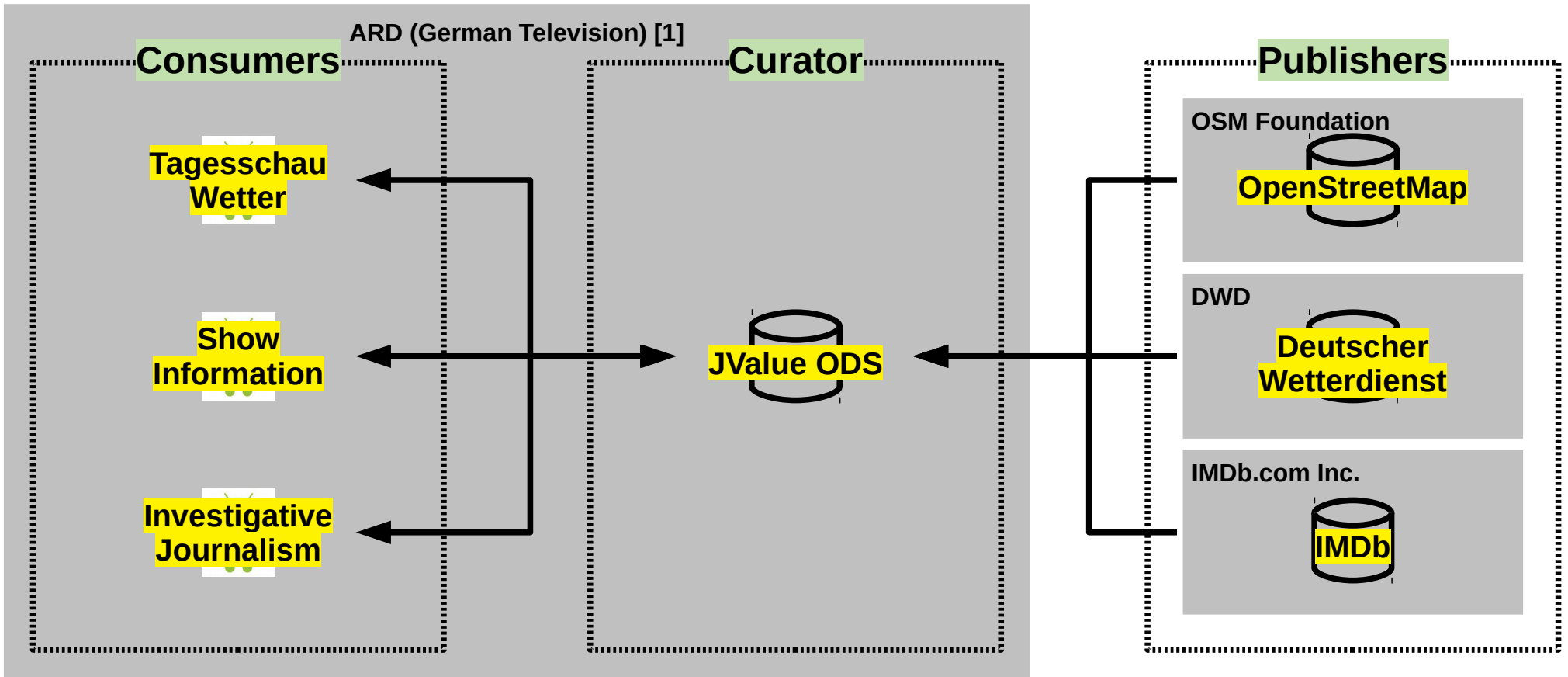
- Decoupling of consumers from curators from publishers ...
  - Prevents software silos, frees data to be open
  - Enables open innovation by third parties
  - Increases overall innovation speed

# JValue ODS Use-Case (Illustrated) 1 / 3



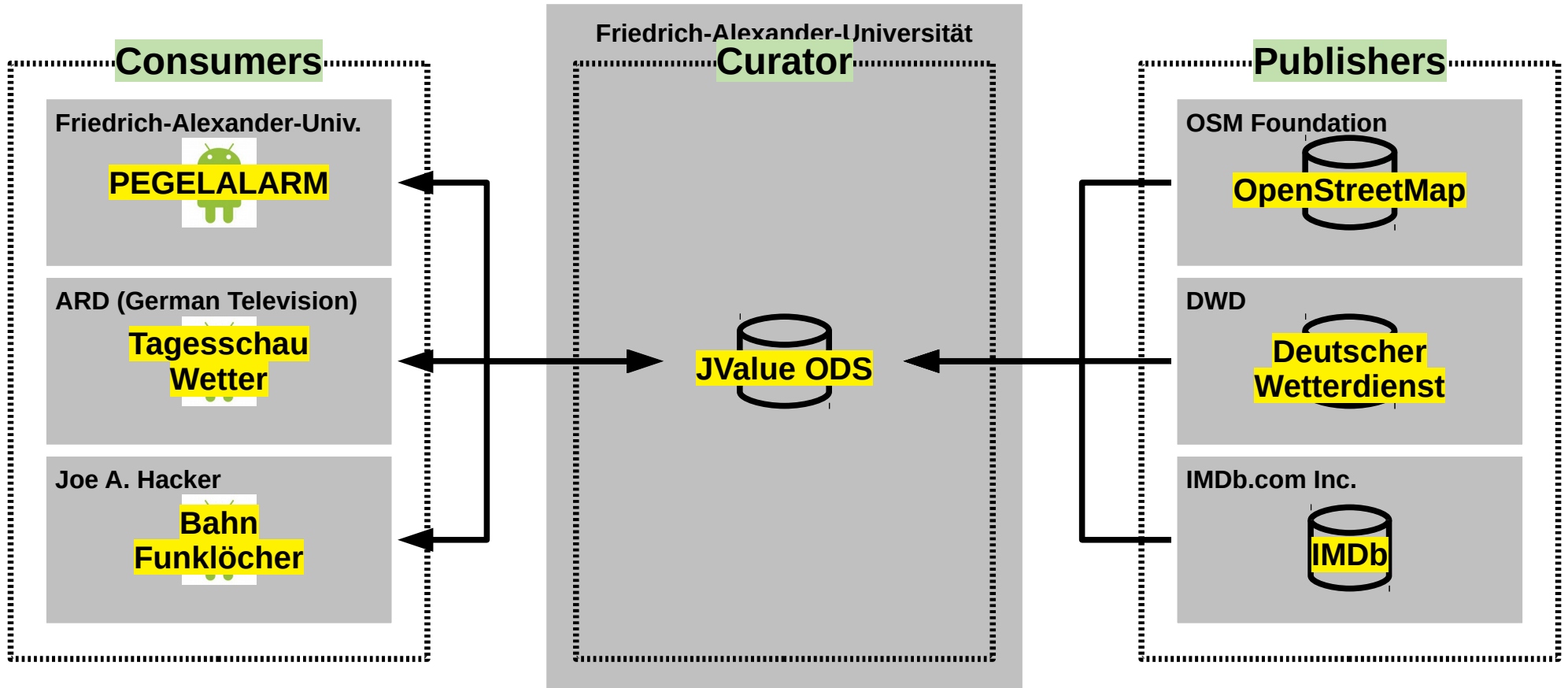
[1] For now, a made-up example; serves illustration purposes only

# JValue ODS Use-Case (Illustrated) 2 / 3

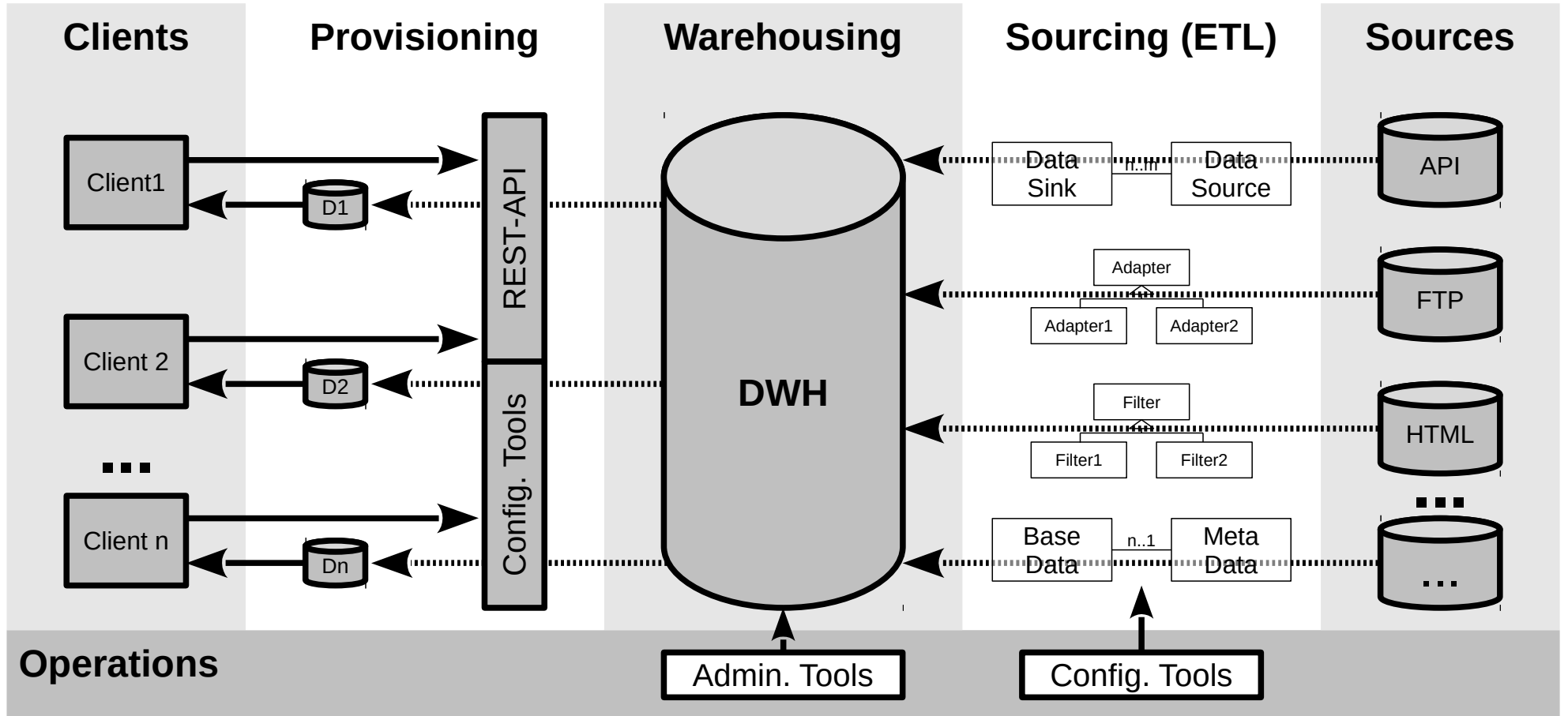


[1] For now, a made-up example; serves illustration purposes only

# JValue ODS Use-Case (Illustrated) 3 / 3



# JValue ODS Architectural Overview



# JValue ODS Base Technologies

- Programming
  - Java and related technologies, REST-API + open source libraries
  - Javascript and related technologies, JSON + open source libraries
- Database
  - Currently a NoSQL database, being replaced with PostgreSQL
  - New microservice architecture will allow variation in databases
- Runtime
  - Runs on Linux and is containerized
  - Deployed as docker images



# Example Open Research Topics 1 / 2

- Software Engineering
  - A reflective REST-API for data discovery
  - Migration to distributed sharded data storage
  - Large-scale cloud operation and management
  - Quota-based operations of data services
  - Real-time event processing
- Data Engineering
  - A data source modeling language
  - A data modeling and manipulation language
  - Data analytics with the JValue ODS

# Example Open Research Topics 2 / 2

- Legal issues
  - Synthetic open data licenses
- Crowd-sourcing
  - Crowd-sourcing open meta-data
- Business
  - Enabling partner ecosystems with open data

# Thanks! Questions?

[dirk.riehle@fau.de](mailto:dirk.riehle@fau.de) – <http://osr.cs.fau.de>

[dirk@riehle.org](mailto:dirk@riehle.org) – <http://dirkriehle.com> – [@dirkriehle](#)