

Survey on: Interoperability and Best Practices in Open Data Portals

Florian Hahn
Data Week 2026
Altes Rathaus Leipzig
June 01-05, 2026

SODS

SURVEY

in DACH
Open data
portals

PAI

Process
Awareness Index
checks for all
collected reponses

Data Week Leipzig 2026

“We publish our data”

The owners of the Open Data portal (technical and administrative) are publishing data into their own ecosystem.

EGovG §12a

**federal
agencies**

have to publish their data

**machine
readiness**

with machine readable
formats

enriched

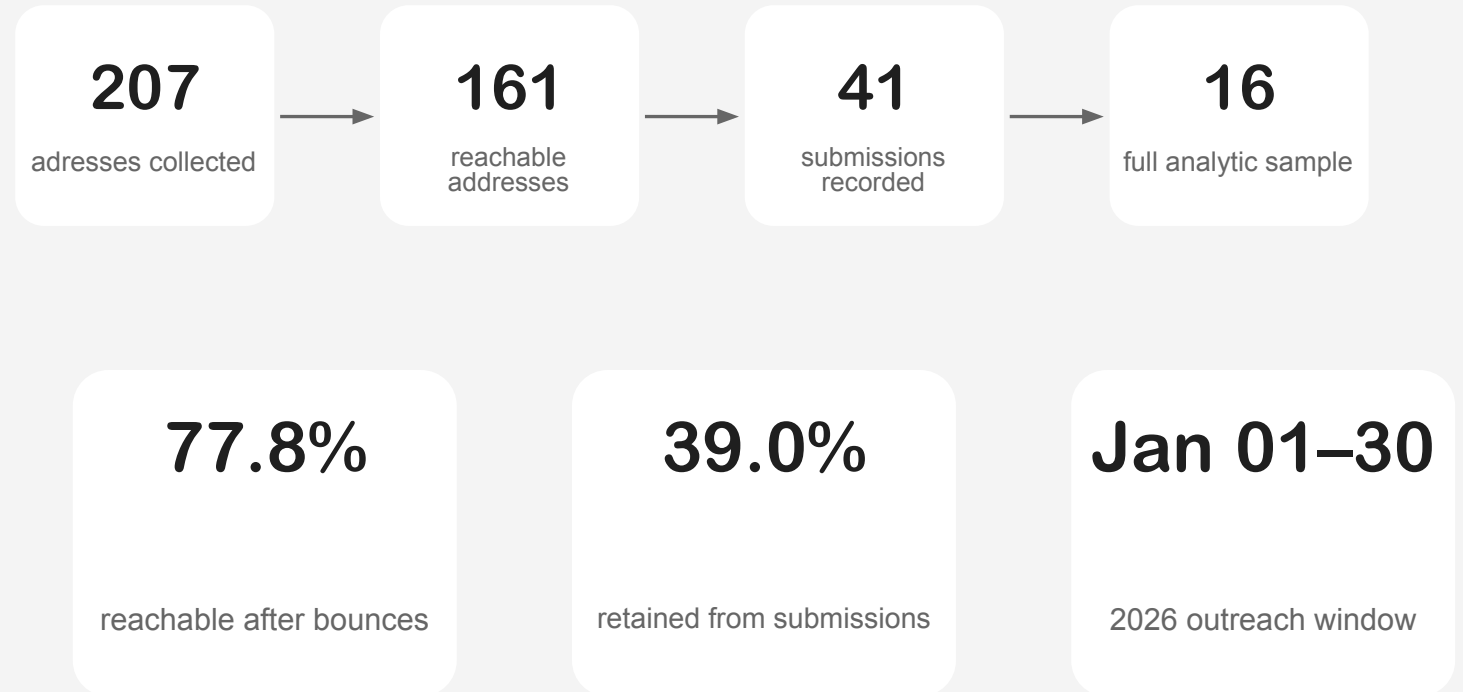
with metadata

None of the above characteristics are clearly defined or technically documented.

From outreach to analytic sample

The study is transparent about deliverability, attrition and the inclusion criterion.

Study design



Inclusion criterion: respondents had to reach at least page 6 (last required page), so core modules were completed.

Interoperability is a process capability

Standards matter, but portal
outcomes depend on routinized
publication, tooling, QA, skills
and organizational capacity.

Motivation

16

analytic survey
responses retained

4

dimensions in the
Process Awareness Index

5/16

above the PAI ≥ 0.60
capability threshold

The issue is not whether one portal publishes data.
The issue is whether many portals publish data in ways that remain comparable.

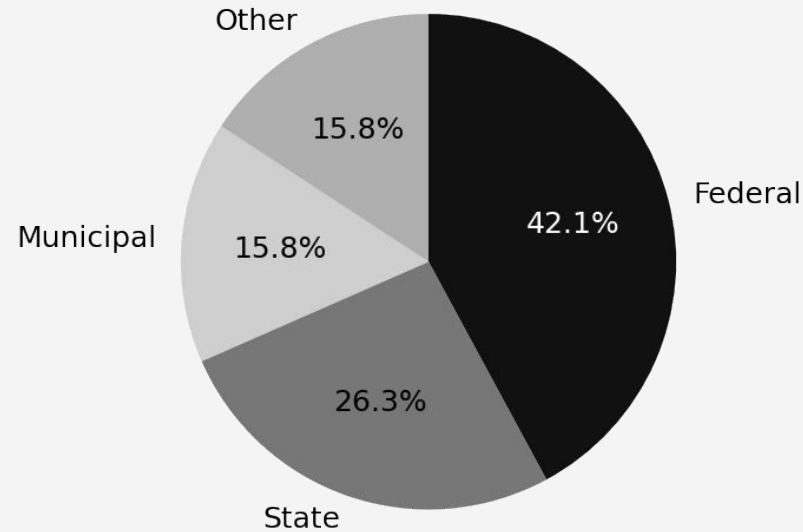
Survey target: operators and stakeholders in Germany, Austria and
Switzerland.

A distributed portal ecosystem

DACH open data is governed across federal, state, municipal and mixed organizational layers.

Setting

Governance level (mentions)



42.1%

federal-level mentions

26.3%

state-level mentions

15.8%

municipal mentions

Governance levels are non-exclusive; at least one respondent selected multiple levels.

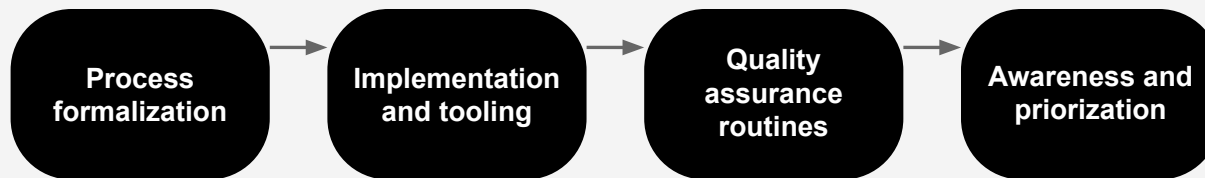
What exactly do we measure?

We ask for the current level of interoperability awareness and implementation among DACH portal operators.

What is the current level of interoperability awareness and implementation among open data portal operators in Germany and the wider DACH region?

Readiness is limited, and many operators have not yet institutionalized interoperability-enhancing publishing practices.

4 dimensions:



Operationalized as a descriptive Process Awareness Index, not as a certification score.

4 dimensions, example questions

All four dimensions have reproducible Survey Codes for the question items, see GitHub repository

Questions examples

Process formalization

Is there a defined process for publishing new data?
How frequently do you typically update datasets?
Who creates the file names and category names?

Implementation and tooling

Which data formats do you provide in your portal according to Tim Berners-Lee's 5-star scheme?
Does your portal provide an application programming interface (API) for accessing the data?)

Quality assurance routines

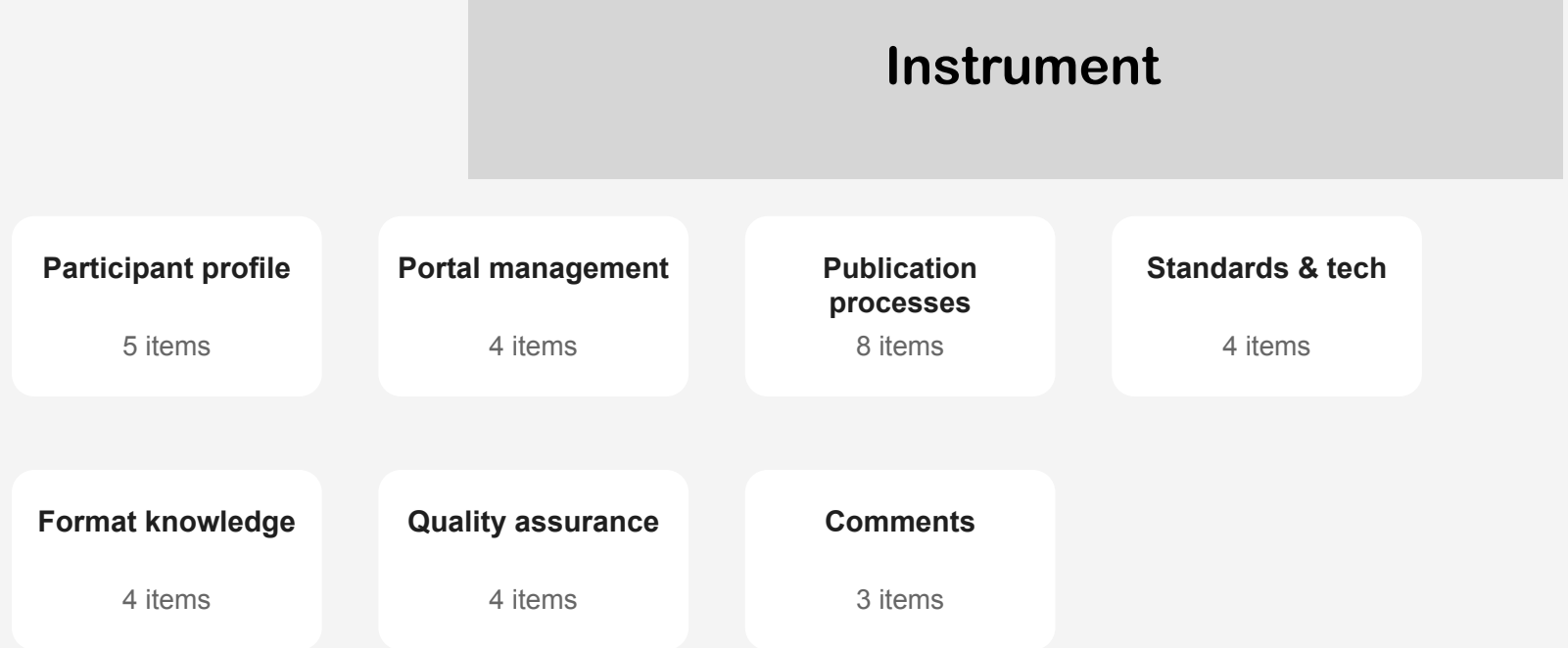
Are there procedures for quality assurance of the data provided?
Do you check your datasets, including metadata, for compliance with best practices (formats, character sets, naming conventions)?

Awareness and prioritization

I am familiar with the DCAT-AP metadata standard.
I am familiar with the principles of Linked Data and RDF.
I am familiar with best practices for working with CSV files compared with other formats (e.g., JSON or RDF).

32 items, 7 modules

The questionnaire links governance, publication processes, standards, data formats, QA and comments.



Question types: single choice, multiple choice, ordinal scales and optional free-text fields.

Designed to reduce vague self-reports by asking behaviorally specific questions.

A descriptive capability index

PAI aggregates four dimensions into a [0, 1] indicator for monitoring and capacity-building.

PAI construction

D1

Process formalization
mean 0.559

D2

Implementation and tooling
mean 0.567

D3

Quality assurance
mean 0.469

D4

Awareness & prioritization
mean 0.525

$$\text{PAI} = \text{mean}(\text{D1}, \text{D2}, \text{D3}, \text{D4})$$

0.60

above-baseline threshold

5/16

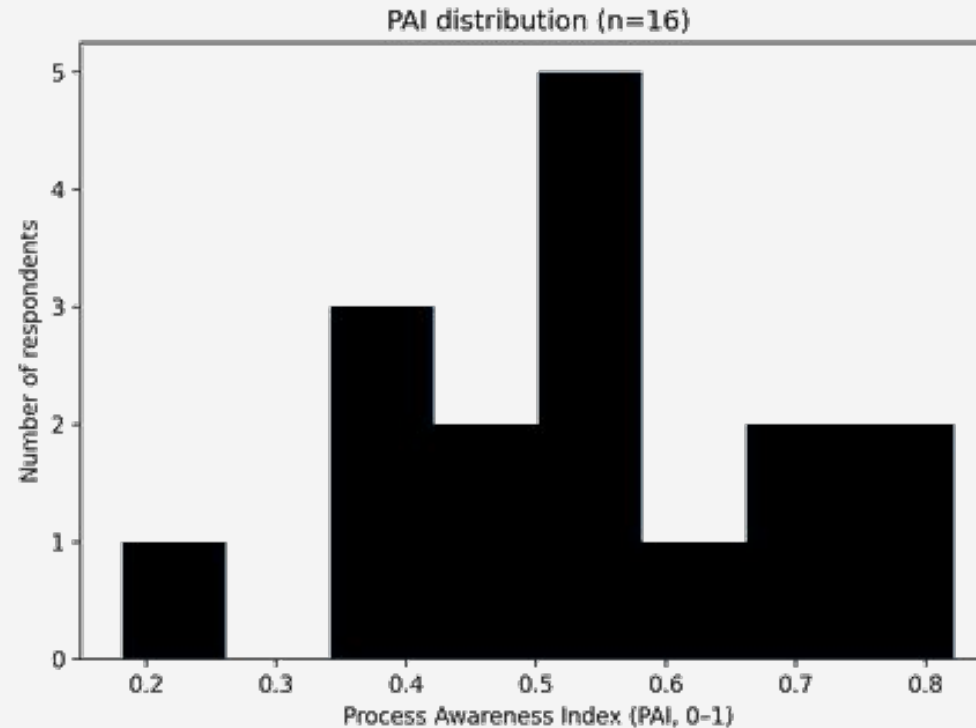
respondents meet it

“Not sure” is treated as an explicit uncertainty signal rather than silent missingness.

Main result

Readiness is moderate but dispersed

The median PAI is 0.521, but the range from 0.203 to 0.762 shows heterogeneous capability profiles.



0.530

mean PAI

0.521

median PAI

0.203–0.762

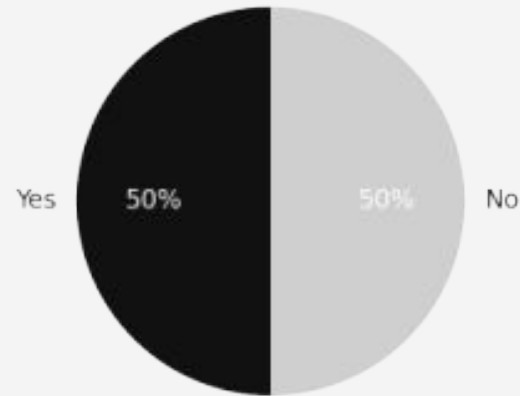
observed range

Interpretation: the sample contains capable portals, but above-baseline readiness is not the majority case.

Fixed publishing processes split the sample

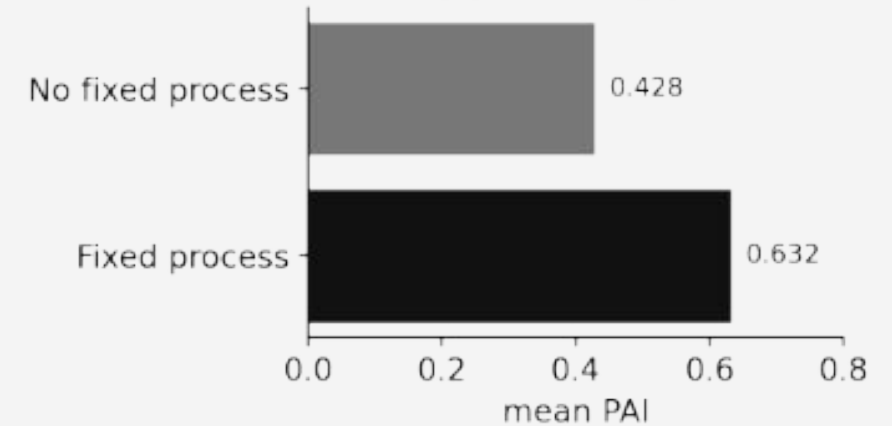
Only half report a defined publication process, and PAI is much higher when a process exists.

Defined publishing process



Process gap

PAI by publishing process



8/16

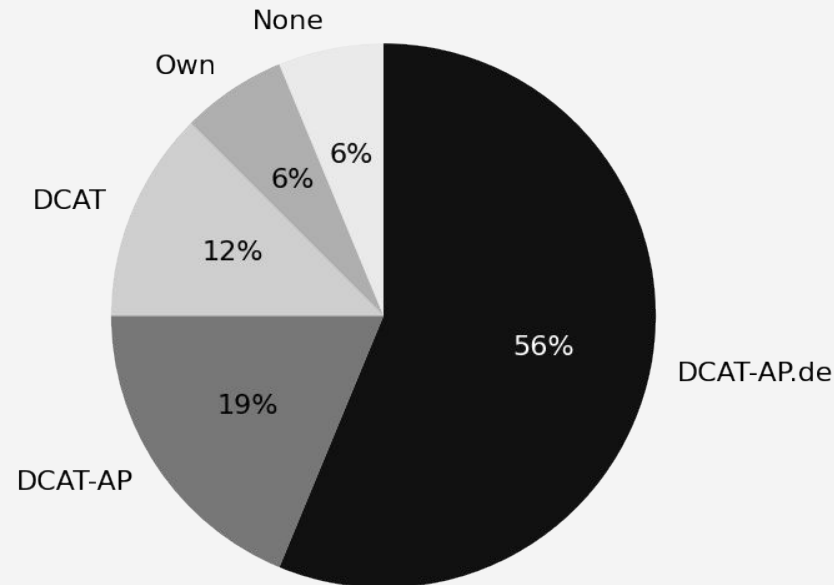
report a fixed process

A process is not bureaucracy here. It is an interoperability control point.

Standards adoption is uneven

DCAT-AP.de is the most frequent profile, but the reported standards landscape is still mixed.

Metadata standards (n=16)



Standards

9/16

DCAT-AP.de

3/16

DCAT-AP

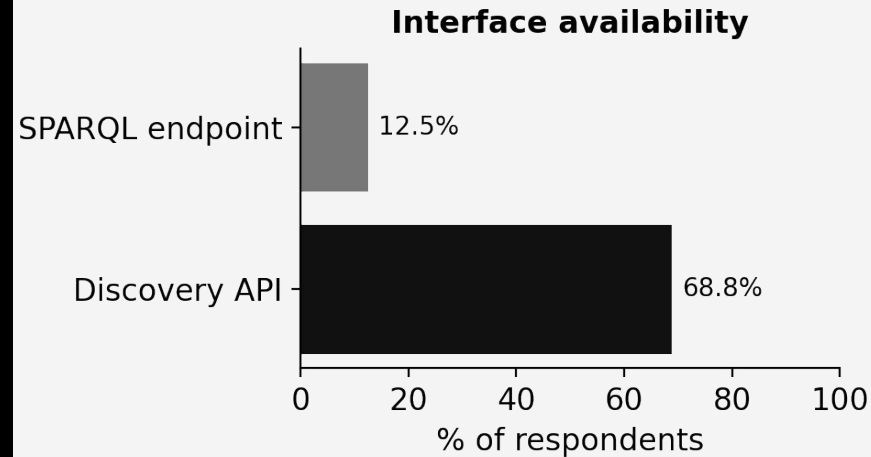
2/16

DCAT

Standard use alone does not guarantee interoperability; implementation routines decide whether profiles become operational practice.

APIs are common, SPARQL is rare

Discovery APIs are reported by a majority, but linked-data query infrastructure remains exceptional.

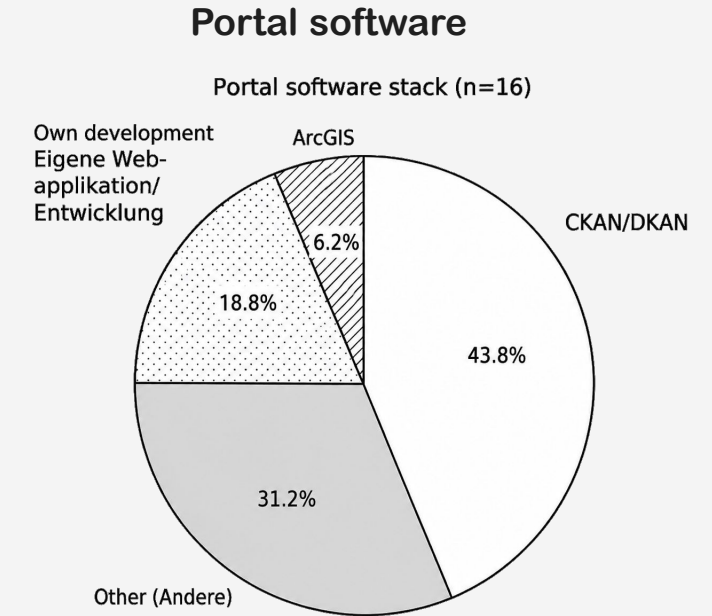


7/16

CKAN/DKAN

5/16

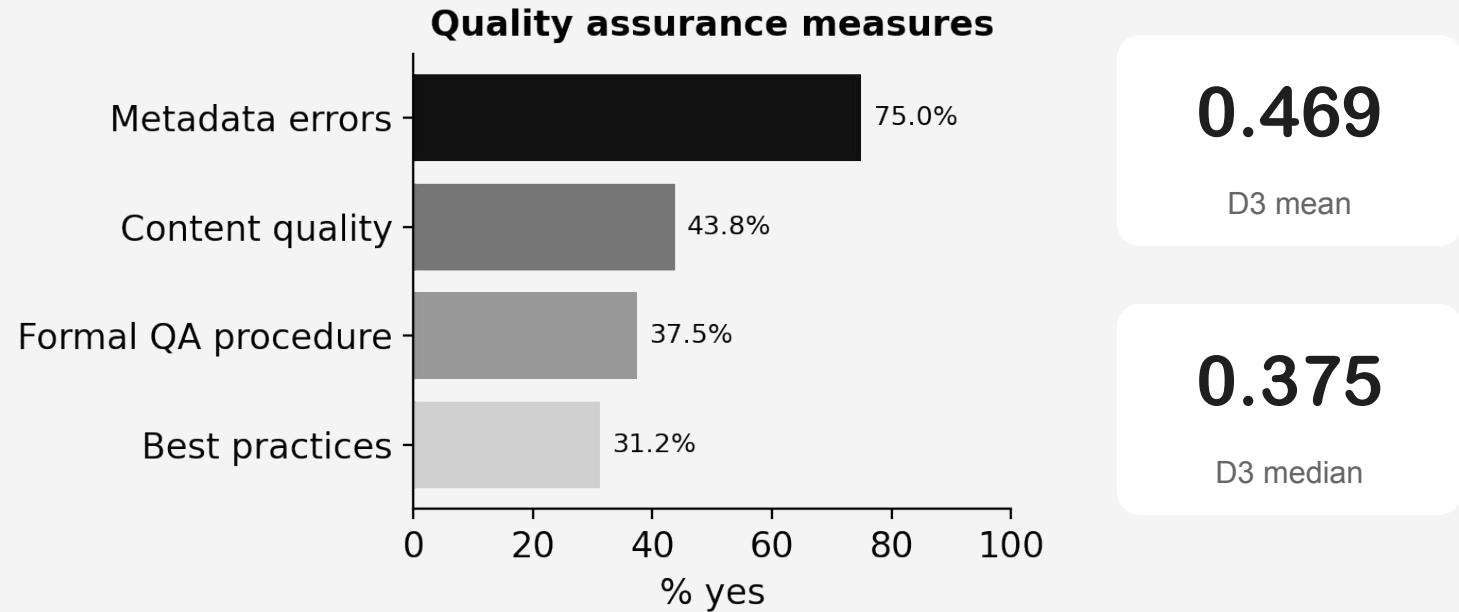
other stack



Tooling heterogeneity is expected in federal settings. The interoperability question is whether common interfaces and validation routines bridge that heterogeneity.

QA is the weakest dimension

Metadata checks are more common than formal data-quality and best-practice routines.



This is the most actionable bottleneck: checklists, validators and lightweight routines can target it directly.

Barriers are mainly capability-related

Optional free-text responses point to knowledge and resource limitations.

Implementation barriers

7
(43,7%)

mentions of missing data
knowledge

5
(31,25%)

mentions of missing domain
knowledge

4
(25%)

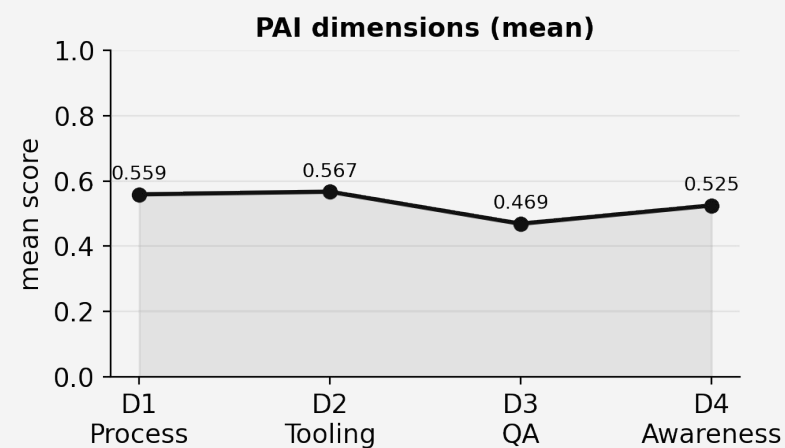
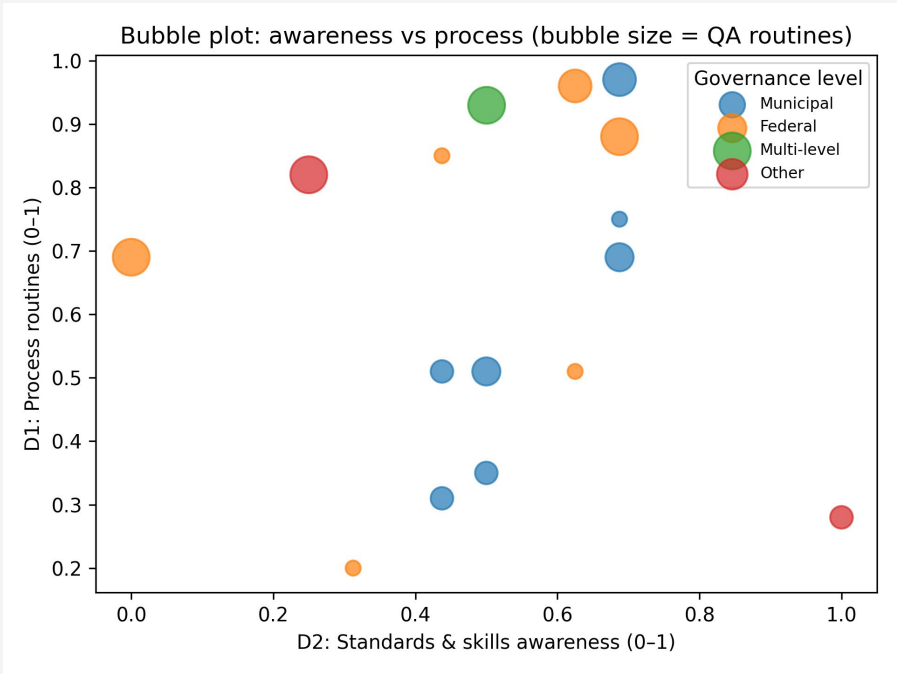
mentions of limited
personnel resources

Do not read these as prevalence estimates. Read them as design clues for interventions.

Capacity building should reduce perceived complexity and translate standards into daily publication routines.

Readiness is not one-dimensional

The bubble plot shows that awareness, process routines and QA do not move uniformly across respondents.



D1
Process formalization

D2
Implementation and tooling

D3
bubble size: QA

D4
Awareness and prioritization

The policy lesson: a single training format will not fit all portals.

From measurement to intervention

SODS turns an abstract interoperability problem into measurable capacity-building targets.

What follows?

1 Training

DCAT-AP constraints, authoritative vocabularies, validation basics

2 Workflow templates

default roles, checklists, update cadence, exception handling

3 Peer learning

translate standard updates into actionable guidance

4 Artifact checks

compare reported routines with published metadata quality

Interoperability readiness becomes a target for program design and longitudinal evaluation.

Open instrument, data and scoring

The repository contains the anonymized processed dataset, scoring specification, survey structure and figures.



github.com/SODIC-research/SODS

Reproducibility

Repository asset	Purpose
LimeSurvey .lss	full instrument and routing
processed CSV	cleaned variables + PAI components
scored results	replication of reported indicators
visualizations	figures used for reporting

The paper also reports a positive institutional ethics decision for the non-identifying research data release.

What SODS does not claim

The study is deliberately conservative: PAI is a capability indicator, not an audit score.

Limitations

N = 16

small specialized sample

self-report

not direct metadata
verification

descriptive

not causal inference

Next step

Pair survey responses with artifact-based checks, for example DCAT-AP validation and metadata quality indicators.

Trainings and White-Papers on Best Practices.

This triangulates perceived readiness with observable publishing practice.

The main message

Interoperability readiness is measurable, uneven and actionable.

Take away

1

Measure capability

PAI makes process awareness visible across routines, tooling, QA and prioritization.

2

Target the bottleneck

QA and fixed publishing processes are concrete intervention points.

3

Monitor over time

The open instrument enables replication and longitudinal readiness tracking.

Open data interoperability fails silently when processes are not institutionalized.

Thanks!

Do you have any questions?

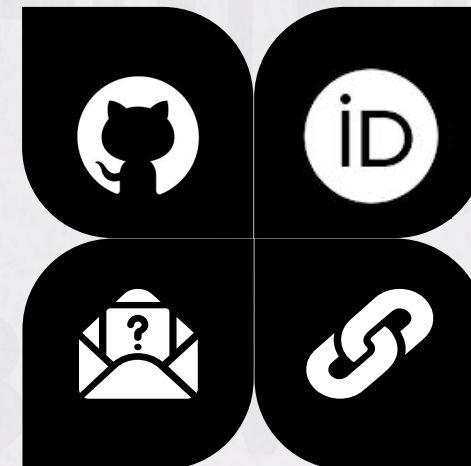


florian.hahn@informatik.tu-chemnitz.de



<https://github.com/SODIC-research>

<https://orcid.org/0009-0008-1126-9319>



florian.hahn@informatik.tu-chemnitz.de

<https://www.tu-chemnitz.de/cs/dm/team/fh.php.en>