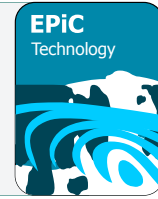




EPiC Series in Technology

Volume 5, 2026, Pages 36–53

Proceedings of 17th International Conference on Current Research Information Systems



The First NRW State Research Report: Role, Challenges and Solutions from the Data Trustee CRIS.NRW

Malte Kramer and Eva Maria Frechen
University of Muenster, Schlossplatz 2, 48149 Münster

Abstract

In Germany's federal system (Bundesstaat), higher education is primarily a responsibility of the 16 federal states (Länder). North Rhine-Westphalia (NRW), Germany's most populous state (~18 million inhabitants), published its first State Research Report in 2025, jointly produced by the Ministry of Culture and Science (MKW) and 32 higher-education institutions (14 universities, 18 universities of applied sciences). CRIS.NRW, a state-funded initiative within the Digitale Hochschule NRW cooperation framework, acted as data trustee. The report is aligned with version 1.3 of the KDSF standard for research information in Germany (Kerndatensatz Forschung), avoids institution-level rankings, and combines aggregated indicators (Part I) with institution-authored narratives in selected KDSF research fields (Part II). This paper presents an in-depth practitioner case study of this first-of-its-kind effort. We address three research questions: (RQ1) How can a data-trustee model reconcile institutional data sovereignty with state-level aggregate reporting? (RQ2) What are the critical data-quality and schema-design challenges when operationalizing KDSF for cross-institutional aggregation? (RQ3) What governance mechanisms and technical artifacts enable repeatable, responsible reporting across heterogeneous CRIS/ERP landscapes? We processed 52 password-protected workbooks comprising 397 validated sheets and 734,328 rows across three reporting years. The paper details governance, pragmatic KDSF profiling, an auditable QA pipeline, the dual perspective of providers and trustee, and a case vignette on awards. We derive five design principles for federated research reporting and conclude with a blueprint for schema-based submissions, shared code lists, and an automation roadmap.

Keywords: Open research information infrastructures; Data trustee governance; Interoperability (KDSF, CERIF, PIDs); Quality assurance and schema validation; Responsible research evaluation; State-level CRIS reporting

1 Introduction

1.1 Context

Germany’s research and higher education landscape is organized federally: the states (Länder) hold primary responsibility for governance, funding and reporting. This creates considerable institutional diversity in systems, definitions and maturity, while public administration, institutional leaders, media, civil society and citizens still require comparable and transparent evidence. NRW, the most populous state (~18 million residents), constitutes one of Europe’s largest state-scale research ecosystems. In this context, the first State Research Report (Landesforschungsbericht, LFB) had to build a reusable evidence base without a historical state-wide baseline.

The mandate framing the LFB defined its objective as a state-wide harmonization of research reporting across public higher-education institutions, bringing together the ministry’s information needs — on behalf of the political sphere, the press and the interested public — with the universities’ shared strategic communication interests. From this mandate, two design principles emerged. First, to adopt the Kerndatensatz Forschung (KDSF) as the shared semantics for research information — the standard recommended by the German Council of Science and Humanities (Wissenschaftsrat) for standardizing research information across institutions (Wissenschaftsrat, 2016). Second, to inform without ranking: the report avoids institutional league tables and instead provides sector-level aggregation for universities and universities of applied sciences (UAS), complemented by structured institutional narratives. The cadence was set to every three years, each issue covering the preceding three years.

1.2 Research questions

This paper reports on a practitioner case study (Yin, 2014) of the first NRW State Research Report and addresses three research questions:

- **RQ1 (Governance):** How can a data-trustee model reconcile institutional data sovereignty with state-level aggregate reporting in a federal system?
- **RQ2 (Data quality and schema design):** What are the critical data-quality and schema-design challenges when operationalizing a national standard (KDSF) for cross-institutional aggregation?
- **RQ3 (Repeatability):** What governance mechanisms and technical artifacts enable repeatable, responsible reporting across heterogeneous CRIS/ERP landscapes?

1.3 Contribution

Our contribution is threefold. First, we provide an empirically grounded account of a state-scale data-trustee model with quantified processing evidence (Section 6). Second, we identify and systematize challenges from both the data-provider and trustee perspectives (Section 7), deriving five generalizable design principles (Section 9). Third, we deliver a reusable blueprint for subsequent cycles and other jurisdictions (Section 10). To our knowledge, this is the first KDSF-based, state-level research report in Germany implemented with a formal data-trustee model and a documented, reproducible QA pipeline. The analytical focus of this paper is primarily on Part I (aggregated indicators), where the trustee pipeline, QA logic and quantitative evidence apply most directly. Observations on Part II—the institution-authored narratives in selected KDSF research fields—are reflected separately in Section 11, where we discuss the heterogeneity of first-round submissions, the resulting revision loop with the institutions, and the implications for the second cycle.

2 Background and related work

2.1 Standards for research information

KDSF. The Kerndatensatz Forschung is a national standard for research information in Germany, first specified in 2016 upon recommendation of the German Council of Science and Humanities (Wissenschaftsrat, 2016). It defines six thematic areas — personnel, doctoral training, third-party funding and finances, patents and spin-offs, publications, and research infrastructures — along with aggregation levels (organizational unit, subject, research field). The standard is maintained by the Commission for Research Information in Germany (KFiD), established by the Joint Science Conference (GWK) in 2021. With the transition from version 1.3 to version 2.0, the KFiD initiated a restructuring from an aggregate-data standard to a base-data model (KFiD, 2024). The present report was produced under KDSF 1.3; future cycles will adopt version 2.0.

CERIF. At the European level, the Common European Research Information Format (CERIF) provides a comprehensive information model for the research domain, maintained by euroCRIS (Jörg et al., 2012). CERIF supports entity-relationship modelling of projects, persons, organizations and outputs, and underpins the OpenAIRE Guidelines for CRIS Managers. While KDSF and CERIF serve different governance contexts — KDSF as a German reporting standard, CERIF as an interchange model — both address the need for semantic interoperability. Azeroual et al. (2019a) have examined mapping and interoperability challenges between the two standards.

FAIR and PIDs. KDSF aligns with the FAIR principles’ emphasis on interoperable, machine-actionable data (Wilkinson et al., 2016). Persistent identifiers support entity resolution: ORCID for person disambiguation (Haak et al., 2012), DOIs for research outputs (Paskin, 2010), and ROR for organizational affiliation (Priem et al., 2022). Beyond these widely adopted identifiers, funder project identifiers (Förderkennzeichen, e.g., DFG project numbers) play a critical role in deduplicating collaborative projects reported by multiple participating institutions, and emerging systems such as the Offenes Register für Auszeichnungen und Preise (ORAP), an open register of awards and prizes published by the KFiD in January 2026 (KFiD, 2026), now support standardization in the awards domain.

2.2 Data quality in CRIS contexts

Quality-assured integration across institutions depends on robust data-quality management (Pipino et al., 2002) and record-linkage techniques (Christen, 2012). Domain-specific work has identified recurring quality issues in CRIS data — including missing values, formatting errors, inconsistencies and duplicate entries — across institutions in multiple countries (Azeroual & Schöpfel, 2019). Azeroual et al. (2018) demonstrated data profiling as a technique for identifying such issues systematically. These findings directly informed our expectations of quality challenges at state scale.

2.3 Privacy and governance

Even in aggregate-only reporting, privacy safeguards are relevant: small-cell policies reduce re-identification risk (Sweeney, 2002). Governance matters: data-trust models — neutral intermediaries working under purpose limitation and collective oversight — can reconcile data sovereignty with shared use in the public interest (Delacroix & Lawrence, 2019).

2.4 Responsible research evaluation

The LFB’s non-ranking design aligns with multiple frameworks for responsible research evaluation: the Leiden Manifesto emphasizes method transparency, context and avoidance of perverse incentives

(Hicks et al., 2015); the San Francisco Declaration on Research Assessment (DORA) recommends evaluating research on its merits rather than via proxy metrics (DORA, 2013); and the Coalition for Advancing Research Assessment (CoARA) Agreement sets a shared direction for assessment reform, emphasizing qualitative judgment supported by responsible use of quantitative indicators (CoARA, 2022). Reporting infrastructures that publish contextualized, aggregate indicators without league tables support these principles.

2.5 Comparable systems

Several national and sub-national research information systems operate with similar building blocks. Norway’s CRISTin (now transitioning to NVA) functions as a centralized national system that documents all scholarly publications by Norwegian researchers, owned by the Ministry of Education and Research (CRISTin, n.d.). This centralized model contrasts with Germany’s federated approach, where institutions remain data owners and share information on request. In Germany, the Forschungsportal Sachsen-Anhalt represents a smaller-scale, manually curated state portal; due to the much smaller number of institutions, however, its challenges in coordination and data heterogeneity differ considerably from NRW’s. These comparisons inform NRW’s positioning: a large-scale, federated, non-ranking model operating between institutional autonomy and centralized infrastructure.

Table 1: Comparison of selected research information systems

Feature	NRW LFB	CRISTin/NVA (NO)	Forschungsportal SA (DE)
Scope	State (32 institutions)	National (~270 inst.)	State (~10 institutions)
Governance	Trustee, federated	Centralized, national	State-coordinated
Semantics	KDSF 1.3	National model	KDSF-oriented
Update cycle	Triennial report	Continuous	Continuous portal
Ranking	Explicitly none	Indirect (pub. levels)	None
Data ownership	Institutional	National database	Institutional

3 Methodology

This paper follows a single-case study design (Yin, 2014), where the case is the creation of the first NRW State Research Report (2021–2023 data, published 2025). The unit of analysis is the end-to-end process from mandate to publication, encompassing governance design, KDSF profiling, data collection, QA, aggregation and reporting. Evidence was collected from four sources: (1) project documents (templates, schemas, meeting minutes, QA logs); (2) processing artifacts (validation scripts, crosswalks, code lists); (3) quantitative process metrics (file counts, row counts, error classifications, iteration counts); and (4) reflective accounts from the authors, who served as the trustee team.

We acknowledge the dual role of the authors as both practitioners and researchers (observer bias). We mitigate this through quantified process evidence, documented artifacts and triangulation with provider feedback from QA round-trips. Limitations of generalizability inherent to single-case designs are discussed in Section 11.

4 Governance and organization (RQ1)

4.1 Roles and mandate

The rectors' conferences (LRK) of universities and UAS mandated a joint working group comprising institutional leaders, domain experts (research administration, libraries, IT) and representatives from MKW. This group agreed scope, structure, cadence and methodological choices. The ministry financed editorial design and published the report (German and English) but had no access to raw institutional data.

CRIS.NRW is a state-funded initiative within the Digitale Hochschule NRW (DH.NRW) cooperation framework, with consortium leadership at the University of Münster and funding from the MKW. Its broader mandate is to support NRW's higher-education institutions in building and integrating current research information systems. In the LFB context, CRIS.NRW served as data trustee under commissioned-processing contracts with defined technical and organizational measures. The trustee's remit was to: integrate institutional submissions; validate and normalize data against a state KDSF profile; deduplicate and resolve entities; compute aggregations; produce QA feedback; and draft method texts. Publication followed a layered model: Part I indicators were published aggregate-only at sector level (universities and UAS, with no institution-level values); Part II consisted of institutional narratives authored by each institution and published under its name; and named individuals appeared in the report only as a documented exception for awards, after shortlist selection and explicit individual consent (Sections 4.3 and 8).

4.2 Trust building

Trust was built on three mechanisms: (1) clear separation of duties — the ministry never saw raw data, the trustee never made editorial decisions; (2) transparent indicator governance — a “no surprises” principle, change logs and a documented sign-off workflow; (3) decision efficiency — decisions were taken through the LRKs rather than in large plenaries, keeping the process tractable while maintaining representativeness. The working group explicitly agreed to avoid institution-level rankings, to adhere to harmonized visual styles, and to disallow institution-specific custom graphics in Part II.

The first iteration presented a particular trust challenge: since institutions could not preview the final report format during data collection, they had to rely on process commitments rather than tangible precedent. Trust in the process — reinforced by the no-ranking guarantee and the sign-off workflow — substituted for trust in a known product. This required considerable willingness from all participants to engage with an open-ended process, and it placed a premium on transparent, continuous communication.

4.3 Legal and ethical safeguards

Aggregate-only publication was enforced as the default, with one documented exception for named individuals in awards (detailed below). A small-cell policy mitigated re-identification risk (Sweeney, 2002). Where person-level checks were technically necessary for QA (e.g., deduplication), pseudonymization and separation of duties applied; access was role-limited; deletion/retention schedules governed processing artifacts. For awards, personal names were shared with the ministry only for a limited shortlist and only after explicit individual consent. These controls align with purpose limitation and data minimization under the commissioned-processing framework.

5 KDSF profiling and report design (RQ2)

5.1 State KDSF profile

KDSF served as the semantic backbone, operationalized via a state profile defining indicator definitions, cardinalities, mandatory fields and permissible aggregation levels. The first hurdle in the working group was not identifying which content areas should appear in the report, but rather establishing the concrete attributes and the appropriate aggregation level for each indicator.

By explicit agreement with the working group, the indicator set was pragmatically extended to include NRW's own state funding lines, even though first-party funding lies outside the strict KDSF project scope. This ensured visibility of state programmes alongside third-party funding.

The profile addressed typical pain points: publication typologies (books, series, proceedings; hybrid cases), third-party funding categories and project roles (lead/partner), patent status and reference-year logic, affiliations/organizational units, and time windows. Crosswalks from local data models to the state profile were documented with validation rules and examples. Where KDSF field aggregation by research fields was not yet available across all institutions, DFG/Destatis subject groupings served as a transitional layer.

5.2 Counting rules (selected)

Key counting rules derived from provider Q&A:

- **DFG structures.** Report collaborative research centres (SFBs), research training groups (GRKs), research units and priority programmes as one project each; capture consortia roles separately where possible.
- **Multi-funder projects.** Where amounts are tracked per funder, it is permissible to split a project into multiple rows with clear identifiers for recombination.
- **Revenue timing.** Report income strictly as per the award decision; do not report post-period adjustments.
- **Research equipment.** Include research equipment grants as third-party funded projects where research-related.
- **Research-related scope.** Include DH.NRW and industry projects only if research-related; exclude non-research development or purely administrative projects.
- **Donations.** Treat donations as third-party funds per Destatis; foundations count under “other funders.”
- **Clinical operations.** Exclude hospital/clinical operations from finance; separate teaching/research from clinics as required by mandated cost accounting.
- **Patents.** List each patent number separately; count EP unitary patents as single patents; mark sold patents as “discontinued”; optionally use IPC/CPC codes for additional aggregation.
- **Awards and memberships.** Record memberships at admission year only (one-off); date awards by announcement year; exclude discontinued prizes; treat state-level prizes explicitly; include personal names only after shortlist selection and individual consent.
- **Programme fields.** Programme-level aggregation (DFG EXC/SFB; EU ERC/MSCA/EIC) was aspirational in round one but infeasible due to missing/heterogeneous labels; future cycles will mandate a programme field with versioned code lists.

Many of these rules could not have been specified in advance. They emerged only after inspecting real submissions — when it became clear, for example, that the unit of a “project” differed fundamentally between institutions (project vs. subproject vs. cost centre), or that consortium roles were handled inconsistently. This underscores the need for pilot runs and iterative tightening (see Principle 4, Section 9).

5.3 Two-part report structure

Part I provides a system overview via aggregated indicators for universities and for UAS, with no institution-level values. Topics include finances (income and expenditure by categories), personnel (by types and subject/discipline groupings), third-party funding (income and projects by funders), the state’s new funding lines, doctoral and habilitation completions, publications (KDSF types), awards/memberships, patents, and gender distributions.

Part II offers each institution a concise, structured narrative for up to three jointly pre-agreed KDSF research fields (sustainability, AI/big data, energy conversion). Each institution was allocated a fixed text template of two pages per research field, ensuring overall report manageability across 30+ institutions while providing sufficient space for meaningful narratives. Several considerations shaped this design: the overall report size needed to remain manageable; institutions needed enough space for genuine self-presentation; and the heterogeneous research landscape of NRW meant that not every institution could provide extensive descriptions for every field. Links and QR codes were permissible; custom graphic styles were not allowed to preserve cross-institutional comparability.

6 Data processing and quantitative evidence (RQ2, RQ3)

6.1 Sources and formats

Data originated from local CRIS systems (projects, outputs, patents), ERP systems (financials and projects), HR systems (personnel) and press-office (PR) sources (awards/memberships). In the first cycle, password-protected spreadsheet workbooks were used as exchange format with templates issued by the trustee. Despite templates, the spreadsheets’ inherent flexibility allowed institutional deviations (additional sheets, columns, formatting), complicating automated ingestion and requiring institution-specific remediation.

6.2 Quantitative overview

Table 2 summarizes the processing evidence across 32 institutions and three reporting years (2021–2023).

Table 2: Quantitative processing evidence

Metric	Value
Institutions	32 (14 universities, 18 UAS)
Workbooks processed	52 (password-protected)
Validated sheets	397
Data rows	734,328
Third-party funding projects (raw)	35,301
Awards/memberships (raw)	2,567
Awards/memberships (curated)	~800 (<31% retention)
Reporting years	2021, 2022, 2023

6.3 QA pipeline

The trustee implemented an auditable, repeatable pipeline:

1. **Export.** Institutions exported data from CRIS/ERP/PR systems using templates.
2. **Secure transfer.** Files were transmitted over protected channels with restricted access.
3. **Schema/rule validation.** Submissions were checked against the state profile; violations were flagged and reported.
4. **Normalization.** Data types, value ranges and references were harmonized; code lists applied (e.g., DFG/Destatis groupings).
5. **Deduplication and entity resolution.** Duplicate detection using PIDs where present; deterministic and probabilistic matching where absent (Christen, 2012).
6. **Aggregation and small-cell protection.** Indicators computed per counting rules; small-cell policies applied with a threshold of $n < 5$, i.e., aggregate cells containing fewer than five contributing entities were suppressed or rolled up into the next available aggregation level (Sweeney, 2002).
7. **QA feedback and round-trips.** Institutions received QA reports highlighting schema violations, outliers and suspected duplicates; corrected submissions were re-ingested.
8. **Sign-off and publication.** Institutions approved relevant aggregates; trustee compiled method texts and passed materials to the publisher.

6.4 QA iteration patterns

Multiple QA round-trips were the rule, not the exception. Approximately 50% of institutions required one additional QA round-trip beyond the initial submission; roughly 40% required two additional rounds; and approximately 10% required three. A notable asymmetry emerged between institution types: universities, due to their larger data volumes, were frequently unable to deliver missing details through manual remediation — the sheer number of records made case-by-case correction infeasible. UAS, with smaller volumes, could more often resolve issues manually. This pattern suggests that the feasibility of manual remediation is inversely related to institutional scale — a finding with direct implications for automation priorities: precisely the institutions with the largest and most complex datasets are least able to compensate for schema gaps through manual effort.

6.5 PID coverage

Persistent identifier coverage was thin in this first iteration. Only a small number of institutions provided PIDs (ORCID, DOI, ROR) for approximately 40% of their records; the majority of institutions did not provide PIDs at all, considering the coverage too low to justify the effort. This confirms findings from CRIS data quality research that PID adoption remains uneven across institutions (Azeroual & Schöpfel, 2019) and underscores the need for institutional PID strategies as a precondition for automated entity resolution at state scale.

Beyond person and output identifiers, funder project identifiers (Förderkennzeichen, e.g. DFG project numbers) are critical for deduplicating collaborative projects reported by multiple participating institutions. When three universities participate in the same collaborative research centre (SFB), the funder project identifier is the only shared key that enables deduplication. Consistent provision of these identifiers was uneven in round one. For awards, the ORAP identifier system was not yet available

during the reporting period; it was published by the KFiD in January 2026 (KFiD, 2026), after the first cycle, and its availability will significantly support standardization in this domain.

6.6 Common error classes

Checks followed core data-quality dimensions — completeness, accuracy, consistency and timeliness (Pipino et al., 2002). Table 3 categorizes the most frequent error types encountered during processing.

Table 3: Frequent error classes

Error class	Examples	Impact
Format heterogeneity	Monetary signs, currency formats, date formats	Normalization effort (days)
Missing mandatory fields	Reference years (patents), funder details	Blocks aggregation
Inconsistent codes	Ad hoc gender codes, PID notations, academic titles	Requires manual harmonization
Heterogeneous free text	Programme labels, award descriptions	Prevents automated aggregation
Scope violations	Non-research projects, non-academic events as awards	Produces low-utility data
Structural deviations	Added sheets, columns, formatting in spreadsheet submissions	Breaks automated ingestion

A practical detail illustrates the effort involved: submission requirements that omitted titles for publications or patents (requesting only subject codes) made it difficult to verify correctness — for instance, to detect whether a leading zero was missing in a subject-area code. Similarly, ad hoc gender encodings (e.g. “wwmmwm” to indicate the gender sequence within a group) and idiosyncratic monetary sign conventions required days of manual validation and correction.

6.7 Preliminary status of official statistics

A practical caveat: in some cases, 2023 figures for finance and personnel were preliminary, pending validation and publication by the statistical offices. The report documents this status transparently in method notes. This is a structural feature of triennial reporting with rolling windows — the most recent year will often carry provisional values.

6.8 Reusable artifacts

The pipeline produced versioned artifacts: a state KDSF profile; crosswalks from local models; validation rules; QA report templates; change logs; and method texts. These artifacts shorten future cycles and support transfer to other jurisdictions. They also serve as institutional learning tools: the crosswalks document how each institution’s local data model maps to the state profile, creating a persistent knowledge base that survives personnel changes.

7 Dual perspective on challenges (RQ2)

7.1 Data-provider perspective

Three domains required paradigm shifts from established reporting routines:

Third-party funding. The LFB required a project perspective with aggregation by subject domains, funders and ideally programme lines — a fundamental departure from ERP-based monetary aggregates. Projects are often split across multiple cost centres in accounting systems, while smaller projects are pooled into collective accounts. This accounting logic does not correspond to a research-project logic. Without CRIS-backed project registers integrating financial, administrative and scientific facets, compilation required extensive manual effort; counting rules diverged across institutions; and mapping to subject groupings and programmes was burdensome. Larger universities were often unable to implement programme-level aggregation comprehensively at scale, and consequently delivered what their existing systems supported. Data quality suffered particularly in this domain.

Awards and memberships. Typically managed as PR lists rather than structured data, awards reporting was entirely novel for the institutions. Press offices maintain lists of prestigious awards for communication purposes, but these are not structured, not systematically collected, and not designed for statistical reporting. Submissions consequently mixed genuine awards with “quasi-awards” — highly competitive funding formats (e.g. ERC grants) where success is perceived as a form of recognition. Absent shared inclusion criteria and minimal fields, downstream categorization was labour-intensive.

Personnel and finance. Institutions are anchored in official statistics with decades-old reporting routines established through the Hochschulpersonalstatistik. Institutions know what to report, they have established report templates in their HR systems. Deviating from these routines for alternative views proved costly and error-prone: new reports had to be created, misunderstandings of requirements arose, and control mechanisms had to be developed from scratch. Several cycles will be needed before comparable data is reliably produced. This is a normal feature of new statistical instruments — it takes multiple iterations until all participants share the same understanding and deliver comparable data.

7.2 Trustee perspective

Schema discipline. For the first iteration, spreadsheets were a pragmatically sound choice: they required no infrastructure prerequisites, were universally available across all 32 institutions regardless of their CRIS maturity, and allowed rapid template development. However, their inherent flexibility — institutions added sheets, columns, formatting — systematically defeated automated ingestion at scale. This is consistent with research on format heterogeneity as a major source of data quality issues in research information systems (Azeroual et al., 2018). The lesson is not that spreadsheets were wrong for round one, but that they cannot remain the submission format for subsequent cycles.

Purpose-fit acceptance. Rule-driven but purpose-misaligned submissions produced low-utility content: projects without financial values, patents without reference years, and events such as internal staff anniversaries or routine ceremonial appointments reported as exceptional awards. Too often, submission requirements were based on strict rules rather than practical purpose, resulting in thousands of data points that could not be meaningfully used. Aligning acceptance criteria with indicator purposes — and communicating concrete examples of in-scope and out-of-scope submissions — would substantially reduce noise and downstream curation costs.

Aggregation without code lists. Programme-level reporting failed without shared code lists and mandatory fields. It had been a deliberate choice not to prescribe programme categories upfront, to avoid prejudicing the data and to allow categories to emerge from the submissions. This approach did not work: most institutions could not provide programme-level information at all, and among those that did, the free-text labels were too heterogeneous to aggregate or extrapolate. Awards similarly lacked

any reference register, forcing heavy manual curation — with nearly every award entry requiring individual research and verification.

The hen-and-egg dilemma. Strict rules before seeing data risk missing important edge cases; too-loose rules generate unmanageable variety. More detailed specifications would have led some institutions to omit data that they could not deliver in the required detail — meaning the associated challenges would never have surfaced and could not be addressed for a second report. This dilemma is structurally inherent to first-iteration reporting in a federated environment. Pilot runs and incremental tightening offer a pragmatic balance. The first report was, in this sense, deliberately conceived as a minimum viable report.

8 Case vignette: awards and memberships

8.1 From unstructured lists to taxonomy

Institutions reported 2,567 awards/memberships (raw). After deduplication and curation, approximately 800 entries met inclusion criteria (<31% retention). The 69% attrition resulted from a combination of duplicate entries across institutions, submissions that were out of scope (e.g. highly competitive grants reported as awards, non-academic events), and entries lacking minimum required detail. Because the curation process was conducted through manual, case-by-case review — over 1,200 individual records required manual research and verification — without systematic logging of rejection reasons, a precise breakdown of attrition categories is not available. This is itself a lesson for designing auditable curation workflows in future cycles (see Principle 5, Section 9).

The challenge was compounded by the complete absence of prior data. Until this first collection, there was simply no data basis on how many awards and memberships were held by researchers at NRW institutions, nor which granting bodies were involved, nor what level of prestige these awards represented. Institutions were initially asked to report all awards and memberships known to them from the reporting period, leading to the heterogeneous initial dataset.

8.2 Exploratory categorization

In the absence of a reference registry — the ORAP register (KFiD, 2026), then under development at the KFiD, was not yet available during the reporting period — the trustee derived a categorization from the data. This categorization had to be developed exploratively, working pragmatically from the data at hand while maintaining traceability, verifiability and repeatability. Five dimensions emerged:

9. **Purpose and objective:** excellence/achievement, early-career support, lifetime achievement, transfer/innovation, citizen science, gender equality, teaching/science communication.
10. **Disciplinary scope:** mono/inter-/transdisciplinary; SDG linkages; alignment with state research priorities.
11. **Eligible audience:** career stage (students, early-career, established); individual/team/institution; sector (higher education, non-university research, industry, civil society); regional/residency requirements.
12. **Recognized work:** single work vs. corpus; publications; patents/prototypes; transfer achievements; teaching/science communication; open-science contributions; engagement/ethics.
13. **Quality criteria:** peer review; jury/expert judgement; evidence basis (without purely metric fixation; DORA-compatible).

A particular challenge was the boundary between awards and quasi-awards: some institutions reported success in highly competitive funding formats (e.g. ERC grants, Leibniz prizes) as a form of recognition, blurring the line between competitive funding and awards. The trustee ultimately had to make qualitative judgements — a delicate task, since no criteria had been specified in advance, and evaluative judgements arguably belong with the report’s addressees rather than the data processor. This tension between neutral trustee role and necessary curation decisions is a structural issue that must be addressed through clearer upfront criteria in future cycles.

8.3 Outcomes and reusability

The resulting categorization schema is reusable for future cycles and provides a foundation that can be strengthened by shared code lists and the ORAP community register (KFID, 2026). Memberships were recorded at admission year only; awards dated by announcement year; discontinued prizes excluded. Small-cell rules and narratives replaced any ranking. Named individuals appeared in the published report as the documented exception to the otherwise aggregate-only publication principle (Section 4.3): personal names were included only after shortlist selection and explicit individual consent, and only for the awards domain.

9 Design principles for federated research reporting

From the empirical evidence documented in Sections 6–8, we derive five design principles that abstract from the NRW case to generalizable guidance:

Principle 1: Schema-first, not data-first. For a first iteration, low-barrier formats (e.g., spreadsheet templates) may be pragmatically necessary to enable universal participation across institutions with heterogeneous technical maturity. However, strictly enforced schemas must be introduced as early as possible in subsequent cycles. The effort consumed by format heterogeneity in round one (Table 3) demonstrates that submission flexibility translates directly into manual normalization burden. What is not defined will be creatively interpreted.

Principle 2: Code-list sovereignty. Aggregation without shared, versioned code lists is infeasible. The attempt to extract programme categories from institutional submissions post hoc failed because free-text labels were too heterogeneous. Code lists must be owned by a designated authority, versioned with change logs, and distributed with mapping workbooks and concrete examples. This applies to programme fields, subjects, awards, organizations, gender/diversity categories and funding sources.

Principle 3: Purpose-aligned acceptance. Data submissions should be evaluated against indicator purpose, not only formal rules. Formally valid submissions that do not serve the indicator’s analytical purpose (e.g. projects without financial values, patents without reference years) create noise and curation cost without information gain. Concrete examples of in-scope and out-of-scope submissions, communicated as part of the template, reduce noise and curation costs substantially.

Principle 4: Pilot before scale. A representative subset of institutions should pilot submissions before the main cycle. This surfaces definition gaps, format issues and edge cases that no specification can fully anticipate — addressing the hen-and-egg dilemma pragmatically. The first NRW cycle effectively served as its own pilot; future cycles benefit from its artifacts, but new indicator domains should still be piloted.

Principle 5: Auditable curation workflows. Manual curation decisions — especially for domains like awards where no reference standard exists — should be logged with rejection reasons, categorization rationale and sources. Without such logging, institutional learning is lost, process transparency is compromised, and dependence on individual expertise increases. The absence of systematic rejection logging in the awards curation (Section 8.1) illustrates the cost of omitting this step.

10 Blueprint for LFB 2.0

10.1 Target audiences

While the original mandate identified target audiences at a high level (policy-makers, media, the interested public), these were not operationalized into distinct personas with mapped indicator sets, granularity levels and visualization styles. For the next cycle, 3–5 personas should be defined (e.g. policy analyst, rectorate strategist, journalist, citizen) and each mapped to specific indicators and visualization approaches. This alignment should drive decisions on language, level of detail and output format.

10.2 Submission standards

Move from flexible spreadsheets to schema-validated CSV/JSON or institutional APIs. Publish machine-readable schemas (e.g. JSON Schema) with field-level documentation. Build automated ingestion with immediate validation feedback: submissions are checked upon receipt, non-conformant parts are flagged and returned, and only conformant content is accepted into the pipeline. The transition to KDSF 2.0, which restructures the specification as a base-data model (KFiD, 2024), provides an opportunity to align submission schemas directly with the new specification architecture.

10.3 Programme fields, code lists and master data

Mandate a programme field with versioned, shared code lists (DFG EXC/SFB/GRK, research units, priority programmes; EU ERC/MSCA/EIC; EU Missions; federal/state lines). Provide mapping workbooks and examples; publish change logs. Extend code lists for subjects/fields (DFG/Destatis; later KDSF fields), organizations (ROR), awards (leveraging ORAP-PRIZE-ID; KFiD, 2026), units (currency symbols), gender/diversity categories and academic titles.

A prerequisite that precedes API-based submission is the harmonization of master data across institutional systems. Institutions must work with shared master-data registries for external organizations (funders, industry partners), award definitions, and funding-source classifications. Without this upstream harmonization, automated system reports from institutional CRIS will reproduce the heterogeneity currently encountered in spreadsheet submissions. The vision is that data submissions to the LFB become a standard system report from the institutional CRIS — a “press the button” operation rather than a manual compilation exercise.

10.4 Counting rules and minimal fields

Codify counting rules for projects (unit of analysis; subprojects; consortia roles), outputs (publication typologies), patents (status/reference year) and awards (inclusion criteria, dating). Define minimal mandatory fields per domain to support deduplication, QA and aggregation; define recommended fields for richer analysis. Normalize PID notations and provide validation logic (ORCID, DOI, ROR, funder project identifiers).

10.5 QA round-trips and change management

Standardize QA report templates and timelines. Provide a shared Q&A forum and office hours for institutions. Pilot submissions from a representative subset of institutions before the main cycle. Publish release notes for profile changes and code-list updates.

10.6 Automation roadmap

Implement staged automation: spreadsheets → schema-validated CSV/JSON → institutional APIs/system reports. Each stage requires preconditions: PID coverage, code-list adoption, schema discipline, and master-data harmonization. Build provider dashboards showing schema conformance, frequent error classes and trend checks. The long-term goal is that the LFB can be assembled largely automatically from institutional CRIS — though fully automatic production will remain aspirational until CRIS adoption and master-data harmonization reach sufficient maturity across the state.

10.7 State research portal

Develop a portal with interactive, accessible views, versioning and export APIs, complementing the triennial report. A portal offers advantages over a static report: it can be more current, provides more space, enables interactivity, and can serve different target audiences through tailored views. Normalize organizations with ROR (Priem et al., 2022). Preserve privacy safeguards and method transparency. Define acceptance criteria both for the report and for a future portal — including data completeness thresholds, timeliness requirements, and quality gates.

11 Discussion

11.1 Answering the research questions

RQ1 (Governance). The data-trustee model successfully reconciled institutional data sovereignty with public reporting needs. The key mechanisms were: separation of duties (ministry does not see raw data), purpose limitation under commissioned-processing contracts, aggregate-only publication with small-cell policies, and collective governance through the LRKs. The model operationalizes bottom-up data-trust principles (Delacroix & Lawrence, 2019) within existing German higher-education governance structures. The no-ranking guarantee served as a critical trust anchor, enabling institutions to share data without fear of competitive misuse.

RQ2 (Data quality and schema design). The critical challenges were: (a) format heterogeneity stemming from spreadsheet flexibility; (b) absence of shared code lists for programme fields and awards; (c) divergent counting rules reflecting differing local accounting structures; (d) thin PID coverage preventing automated entity resolution; and (e) purpose-misaligned submissions. These findings are consistent with domain-specific CRIS data quality research (Azeroual & Schöpfel, 2019; Azeroual et al., 2018) and extend it to the specific context of federated, state-level aggregation. A distinctive feature of the NRW case is the scale: 32 institutions, three years, over 700,000 rows — magnifying the impact of each individual inconsistency.

RQ3 (Repeatability). The pipeline produced versioned, reusable artifacts (state profile, crosswalks, validation rules, QA templates, code lists). The quantified QA iteration patterns (Section 6.4) and error classification (Table 3) provide concrete baselines for future cycles. The five design principles (Section 9) abstract from the NRW case to generalizable guidance. Critically, the first cycle also produced institutional learning: the crosswalks now document how each institution's local model maps to the state profile, creating persistent knowledge that supports continuity. These artifacts establish reusability in principle; in practice, however, a non-trivial share of the first-cycle workflow, notably award curation, the resolution of consortium-level project counts, and the Part II revision loop, relied on case-by-case manual judgement. This manual share, together with the heterogeneity of submissions documented throughout this paper, bounds the constancy that can realistically be expected when the pipeline is re-run by different teams or in different cycles. Reusability is therefore better described as

substantial but partial: the artifacts shorten the path for a second cycle without rendering it deterministic. We reflect on the Part II side of this caveat in more detail below.

11.2 Responsible evaluation

The LFB's non-ranking stance, method transparency and narrative context align with responsible evaluation principles. The Leiden Manifesto emphasizes that quantitative evaluation should support, not replace, expert assessment (Hicks et al., 2015). DORA recommends evaluating research on its merits (DORA, 2013). The CoARA Agreement calls for assessment practices based on qualitative judgment supported by responsible use of quantitative indicators (CoARA, 2022). The LFB implements these principles at scale by publishing aggregate, contextualized indicators with explicit method documentation and no league tables. Part II's institutional narratives further reinforce this approach by allowing institutions to present their research in context rather than through indicators alone.

11.3 Comparison with centralized models

NRW's federated trustee model stands in contrast to Norway's centralized approach through CRISTin/NVA, where all institutions report into a national database. The centralized model offers stronger data consistency but requires surrendering institutional data control — an approach that would not correspond to the understanding of autonomy prevailing in German higher education (Wissenschaftsrat, 2016). The NRW model demonstrates that federated approaches can produce comparable results through semantic alignment (KDSF profiling) and a neutral trustee, albeit at higher coordination cost. The explicit price of federation is the QA burden documented in this paper; the explicit benefit is institutional ownership and voluntary participation.

11.4 Heterogeneity in Part II and the revision loop

Part II of the report, institution-authored narratives in three jointly pre-agreed KDSF research fields, was deliberately designed with substantial flexibility. Institutions were explicitly permitted in this first cycle to present their research in heterogeneous ways, reflecting the diversity of the NRW landscape and the absence of a shared narrative tradition for state-level reporting. This openness was a conscious trade-off: it lowered the participation barrier and avoided premature standardisation of how institutions describe themselves.

In practice, however, the heterogeneity of first-round Part II submissions exceeded what could be reconciled across the report. The trustee therefore initiated an additional revision loop in which institutions received concrete feedback together with harmonization suggestions, on terminology, comparable level of detail, and a common narrative structure across the three research fields. This iteration produced a substantially more coherent and higher-quality Part II without prescribing content, but it was only feasible because the report was still in production and the editorial timeline allowed for a further QA round trip.

For the second cycle, the experience suggests that institutions will need to coordinate more closely upfront. Pre-agreed narrative scaffolds, examples of well-structured contributions, and an early exchange between institutions on the chosen research fields are likely necessary to enable comparable, high-quality narratives without the burden of a downstream revision loop. This recommendation complements the schema-first orientation for Part I (Principle 1, Section 9): Part II benefits from light-touch harmonization of narrative form rather than rigid schemas of content.

11.5 Generalization

Transfer to other states or national contexts requires: a clear mandate, a neutral trustee and collective governance. The biggest accelerators are schema-validated submissions, versioned code lists and PID

normalization. The principal risks are overreach (scope too large for a first iteration), underspecification (too few rules for comparability), and “spreadsheet drift” (format proliferation).

The trustee model is scalable in both directions. At the state or national level, it serves as intermediary between institutions and public reporting. At the institutional level, the same pattern can be applied with a central unit — such as research services, the library, or an institutional CRIS team — acting as a neutral instance between faculties and university leadership. This internal trustee manages data integration, QA, and reporting while respecting faculty-level data ownership. The design principles (Section 9) apply at both scales.

11.6 The KDSF 1.3 to 2.0 transition

The first LFB was produced under KDSF 1.3. The transition to KDSF 2.0, which restructures the specification from an aggregate-data standard to a base-data model (KFiD, 2024), will shape the next cycle. For the CRIS.NRW consortium, this transition primarily changes the perspective from which data is viewed rather than the way data is handled operationally: CRIS.NRW has worked with base-level data in its institutional CRIS implementations from the outset, and the aggregate indicators required for the LFB were reporting elements derived from that base. The restructuring of KDSF thus aligns the standard with existing CRIS practice, which may lower the barrier for institutions that have already adopted CRIS-based workflows.

11.7 Threats to validity

Construct validity: The process metrics (files, rows, iterations) are objectively recorded; qualitative assessments of challenges draw on documented QA logs, provider correspondence and meeting minutes, reducing subjectivity.

Internal validity: As a descriptive case study, causal claims are limited. We interpret patterns (e.g. the inverse relationship between institutional scale and manual remediation feasibility) as observed associations, not confirmed causal relationships.

External validity: This is a single-case study in one German state. While NRW is the largest state and includes diverse institution types (research universities, UAS, confessional institutions), transferability to states with different governance cultures, smaller scope, or more mature CRIS landscapes requires careful contextualization. The comparison with CRISin and Sachsen-Anhalt (Table 1) provides initial orientation but does not substitute for replication.

Reliability: The documented pipeline, versioned artifacts and quantified process evidence support reproducibility. The dual role of the authors as practitioners and researchers introduces potential observer bias, partially mitigated by triangulation with process data and provider feedback.

11.8 Limitations

The initial mandate identified target audiences at a high level but did not operationalize them into personas with specific indicator requirements, affecting structure and visualization choices. PID coverage remains thin; the majority of institutions did not provide PIDs at all in the first cycle. Programme-level aggregation was infeasible without shared code lists. Awards lacked a reference register; categorization required heavy manual curation without systematic logging of rejection reasons. Alternative views on personnel conflicted with established HR reporting routines and will require multiple cycles to stabilize. Some 2023 official-statistics values were preliminary at the time of drafting. The layout process was highly iterative, dependent on first seeing the data before visualization decisions could be made — a feature inherent to first-of-its-kind reporting. These limitations are addressed in the blueprint for LFB 2.0 (Section 10).

12 Conclusion

The first NRW State Research Report demonstrates that a federal state can assemble a transparent, contextualized picture of its research ecosystem by combining a neutral data-trustee model, a pragmatic KDSF profile and a repeatable QA pipeline. The dual perspective of providers and trustee surfaces five design principles — schema-first submission, code-list sovereignty, purpose-aligned acceptance, pilot-before-scale, and auditable curation — that abstract from the NRW case to generalizable guidance for federated research reporting.

The first cycle was deliberately conceived as a minimum viable report: conservative in scope, transparent about limitations, and designed to create the artifacts and institutional learning that make subsequent cycles more efficient. The quantified effort — 52 files, 397 validated sheets, 734,328 rows, 35,301 projects, 2,567 awards curated to approximately 800 — illustrates both the scale and feasibility of such an undertaking.

The blueprint for LFB 2.0 — including the adoption of KDSF 2.0, schema-first ingestion, programme code lists, master-data harmonization, staged automation and a public portal — charts a path from first-of-its-kind consolidation to a sustainable, interoperable reporting fabric. Within Germany's federal system, NRW's LFB offers a transferable model that respects institutional data sovereignty, avoids rankings, and delivers aggregate, contextualized insights aligned with responsible evaluation principles. The trustee pattern scales both upward (to national contexts) and inward (to institutional reporting), and the CRIS community can adapt it to other jurisdictions, strengthening open research information infrastructures while upholding privacy and responsible evaluation.

13 Acknowledgments

We thank the participating institutions, the rectors' conferences, and the Ministry of Culture and Science of North Rhine-Westphalia for constructive collaboration. Special thanks go to institutional teams who provided data and engaged in multiple QA round-trips. CRIS.NRW is a cooperation initiative within the Digitale Hochschule NRW framework, with consortium leadership at the University of Münster. Trustee work for the LFB was carried out by CRIS.NRW.

References

- Azeroual, O., & Schöpfel, J. (2019). Quality issues of CRIS data: An exploratory investigation with universities from twelve countries. *Publications*, 7(1), 14. <https://doi.org/10.3390/publications7010014>
- Azeroual, O., Saake, G., & Schallehn, E. (2018). Analyzing data quality issues in research information systems via data profiling. *International Journal of Information Management*, 41, 50–56. <https://doi.org/10.1016/j.ijinfomgt.2018.05.007>
- Azeroual, O., Saake, G., Abuosba, M., & Schöpfel, J. (2019a). Solving problems of research information heterogeneity during integration — using the European CERIF and German RCD standards as examples. *Information Services & Use*, 39(1–2), 105–122. <https://doi.org/10.3233/ISU-180030>
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer. <https://doi.org/10.1007/978-3-642-31164-2>
- CoARA (2022). *Agreement on Reforming Research Assessment*. Coalition for Advancing Research Assessment. <https://coara.eu/agreement/the-agreement-full-text/>
- Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts. *International Data Privacy Law*, 9(4), 236–252. <https://doi.org/10.1093/idpl/ipz014>
- DORA (2013). *San Francisco Declaration on Research Assessment*. <https://sfdora.org/read/>

- Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID. *Learned Publishing*, 25(4), 259–264. <https://doi.org/10.1087/20120404>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429–431. <https://doi.org/10.1038/520429a>
- Jörg, B., Dvořák, J., & Jeffery, K. (2012). CERIF: The Common European Research Information Format Model. *Procedia Computer Science*, 33, 257–264.
- KFiD (2024). *KDSF Version 2.0: Restructuring from aggregate-data standard to base-data model*. Kommission für Forschungsinformationen in Deutschland. <https://www.kerndatensatz-forschung.de/>
- KFiD (2026). *Stammdaten – Offenes Register für Auszeichnungen und Preise (ORAP)*. Kommission für Forschungsinformationen in Deutschland. https://kerndatensatz-forschung.de/index.php?id=stammdaten_orap
- Paskin, N. (2010). Digital Object Identifier (DOI) system. In *Encyclopedia of Library and Information Sciences* (3rd ed.), 1586–1592.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211–218. <https://doi.org/10.1145/505248.506010>
- Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*. <https://doi.org/10.48550/arXiv.2205.01833>
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 557–570. <https://doi.org/10.1142/S0218488502001648>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wissenschaftsrat (2016). *Empfehlungen zur Spezifikation des Kerndatensatz Forschung* (Drs. 5066-16). Berlin. <https://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>
- Yin, R. K. (2014). *Case Study Research: Design and Methods* (5th ed.). Sage.