

BEYOND INVENTORY

Why Nobody Used the Data Catalogue

The 2018 metadata rush exposed a deeper failure of operating accountability

Giovanni Leonardi

August 2018



Why Nobody Used the Data Catalogue

Transformation · No. 219

Originally written August 2018 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

"The programme had measured the production of records, not the reduction of uncertainty."

1. The catalogue went live. The questions did not stop.

On the Monday after launch, a business analyst searched the new catalogue for "customer consent". The first result was a table called CUST_MSTR. The second was an extract labelled CRM_OUT_04. Neither entry said whether the data could be used for a retention campaign, which system held the authoritative consent record, or who could decide when two timestamps disagreed.

The analyst did what people had always done: she telephoned the data warehouse manager.

That small scene explains why so many data catalogues are becoming technically complete and practically irrelevant. In the rush surrounding the General Data Protection Regulation, organisations have discovered metadata. They are buying repositories, defining glossaries and asking data owners to attest inventories. The activity is understandable. Article 30 records, subject access requests, retention decisions and consent controls all demand a clearer account of what information exists and how it moves.

But a catalogue that merely describes data is not yet a useful instrument. It becomes useful only when it changes how a person makes a decision under pressure.

My central contention is simple: the data catalogue fails when it is governed as a library and succeeds when it is governed as part of the operating model. The difference is not the richness of the metadata. It is whether the catalogue has authority at the moments when work is assigned, access is granted, a definition is disputed or a control fails.

2. The inventory mistake

The prevailing response to uncertainty has been to collect more descriptions. Project teams create templates with fields for definition, owner, source system, classification, retention period and lawful basis. Completion percentages rise. Steering papers report that 82 per cent of critical data elements have been catalogued.

The number is comforting and often meaningless.

Consider a composite retail bank in early 2018. Its programme identified 1,240 "critical" data elements across finance, risk, customer service and marketing. After six months, 1,017 had entries in the catalogue. The dashboard showed 82 per cent coverage. Yet a sample of 60 entries revealed:

- 41 named a department as owner rather than a person with decision rights.
- 34 repeated a physical column name as the business definition.
- 27 recorded a source system but not the downstream reports or extracts.
- 19 assigned "consent" as the lawful basis without distinguishing the purpose for which consent had been obtained.
- only 8 told a user what to do when the data was missing, late or contradictory.

The programme had measured the production of records, not the reduction of uncertainty. It could answer "Is there an entry?" but not "Can someone act safely because the entry exists?"

This is the inventory mistake: treating the catalogue as the destination of knowledge rather than the route by which a decision reaches the right accountable person.

3. Why people route around the catalogue

When an experienced operator ignores a new repository, the easy explanation is resistance to change. The more useful explanation is that the repository has not earned a place in the workflow.

Three mechanisms recur.

3.1 It is slower than the informal network

A catalogue asks a user to translate a business question into the vocabulary of systems and tables. The informal network does the reverse: a trusted colleague interprets the question, supplies context and accepts some responsibility for the answer.

If "customer" has five definitions, a glossary page listing all five is not enough. The user needs to know which definition governs arrears reporting, which governs marketing suppression and who can resolve a new case. Without that decision route, the catalogue documents ambiguity but does not contain it.

3.2 Ownership is recorded without authority

Many catalogues contain an owner field because the template requires one. The nominated owner may be a senior executive who cannot inspect exceptions, or a subject-matter expert who can diagnose the issue but cannot direct another function to correct it.

Real ownership has three components:

- scope: the data and uses for which the person is answerable;
- decision rights: the questions that person may settle without escalation;

- service expectation: how quickly a user can expect a decision or response.

Remove any one and ownership becomes ceremonial. A name in a field does not create accountability.

3.3 Contribution is separated from consequence

The people asked to populate catalogues are commonly data architects, analysts and project staff. The people who bear the consequence of weak metadata are commonly operations managers, compliance teams, report owners and customers waiting for a correction.

When contributors do not experience the cost of an incomplete entry, completeness becomes a documentation standard rather than an operational necessity. The natural result is technically valid, thin metadata.

A catalogue is not adopted because its contents are comprehensive. It is adopted because the organisation makes important decisions through it.

4. The strongest case for catalogue-first

There is a serious opposing view. An organisation cannot redesign every process before it has an inventory. The estate is too fragmented, ownership is too uncertain and regulatory deadlines are too close. A catalogue-first programme creates a common foundation; usage can follow once coverage is sufficient. Insisting on immediate workflow integration may slow the essential work of discovery and leave the organisation unable to demonstrate reasonable control.

That argument has force. In 2018, many institutions genuinely do not know where personal data is copied, which spreadsheets feed formal returns or how long extracts remain on shared drives. Discovery is not optional, and no responsible practitioner should pretend otherwise.

The flaw lies in making catalogue-first mean catalogue-alone. Discovery and use need not be sequential. A thin entry connected to a live decision is more valuable than a rich entry waiting for a future operating model.

Start, for example, with the data needed to answer a subject access request. Record only what the case team must know: where to search, the responsible contact, expected retrieval time, relevant exclusions, common identity-matching problems and the route for disputes. When a request exposes an omitted archive or ambiguous identifier, improve the entry. Coverage grows through use, and the catalogue acquires credibility because it helps finish real work.

The catalogue-first position is therefore right about the need for a foundation and wrong about how foundations are built. A foundation gains strength from the loads it is designed to carry. Metadata gains strength from the decisions it must support.

5. Replace catalogue coverage with decision coverage

The more revealing management question is not "How much data have we catalogued?" It is "Which recurring decisions can now be made with less delay, less interpretation and clearer accountability?"

A practical portfolio of decision journeys might include:

Decision journey	What the catalogue must reveal	Evidence of usefulness
Approve access to a customer dataset	purpose, classification, owner, restrictions, approval route	fewer referrals and fewer inappropriate approvals
Respond to a subject access request	systems to search, identifiers, retrieval contact, known gaps	shorter search time and fewer missed sources
Resolve a disputed management figure	definition, lineage, transformation rule, report owner	reduced reconciliation effort
Retire an obsolete extract	consumers, retention rule, operational dependency, archive decision	fewer unknown dependencies at change
Assess a new marketing use	source, consent wording, purpose, suppression rule, accountable decision-maker	fewer late compliance escalations

This shift changes both build order and governance.

1. Select a decision journey. Choose a recurring, costly point of uncertainty, not a data domain in the abstract.
2. Observe the work. Follow two or three real cases from question to resolution. Record telephone calls, spreadsheets, hand-offs and quiet judgement calls.
3. Define the minimum decision record. Capture only the metadata needed to complete that journey safely.
4. Name the resolver. Assign the person or forum that can settle ambiguity, with a response expectation.
5. Place the catalogue in the route. Link access approval, change control, incident handling or request management to the relevant entry.
6. Measure the changed outcome. Track elapsed time, referrals, reopened cases and unresolved exceptions.
7. Expand from evidence. Add fields and domains when repeated use demonstrates their value.

This is not an argument for impoverished metadata. It is an argument for sequencing richness behind usefulness.

6. The operating contract behind every entry

A useful catalogue entry is a small operating contract. It says not only what the data means, but how the organisation will behave around it.

For a high-value element, the contract should answer five questions:

- Meaning: What does the term include and exclude in this context?
- Provenance: Where is it created, and what transformations matter to interpretation?
- Permitted use: For what purposes may it be used, under what condition or restriction?
- Accountability: Who decides when meaning, quality or use is contested?
- Remedy: What happens when the data is wrong, unavailable or used outside the agreed purpose?

The fifth question is consistently underrated. Most catalogues describe the steady state. Operations live in the exception state.

Suppose a monthly conduct report shows 18,642 eligible customers while the marketing suppression file shows 18,107. A descriptive catalogue may define both fields correctly. An operational catalogue also states that the report owner raises a reconciliation case, the customer data steward has two working days to classify the difference, and unresolved cases go to the information governance forum before the next campaign selection. The entry becomes valuable at the exact moment certainty breaks down.

The numbers matter here. If the average dispute previously required nine e-mails across twelve working days, and the explicit route reduces it to one logged case resolved in four days, the benefit is not "better metadata". It is eight days of decision latency removed, with an auditable record of who decided what.

7. Governance must move from curation to consequence

Catalogue governance often concentrates on standards: mandatory fields, naming conventions, review dates and quality scores. These disciplines are necessary, but they favour the curator's view of the world.

Operating governance asks different questions:

- Which business decisions rely on this entry?
- What is the consequence if it is wrong?
- Who has used it in the last quarter?
- Which unresolved questions recur?
- When did the accountable owner last exercise a decision right?

These questions expose dormant entries and ceremonial ownership. They also permit proportionate control. A field used only for internal analysis does not need the same treatment as a field that drives customer communication, a regulatory return or a credit decision.

The implication is uncomfortable: some catalogued data should receive less attention. Programmes are reluctant to say this because universal coverage produces a tidy target. Yet equal treatment of unequal data is not rigour. It is an avoidance of judgement.

We should expect a living catalogue to be uneven. High-consequence entries will be richer, reviewed more often and linked to active controls. Low-consequence entries may remain skeletal until a decision journey gives them greater importance. This asymmetry is evidence of management, not failure.

8. What leaders should ask now

The immediate temptation is to commission another completeness campaign. A better intervention is to take the ten most-used catalogue entries and sit beside the people who supposedly use them.

Ask them to complete a real task without calling the person they usually call. Note where the entry fails: missing context, unclear authority, absent lineage, no remedy. Then correct the operating route, not merely the wording.

Leaders should also stop accepting adoption as a count of log-ins. A person may visit the catalogue, fail to find an answer and return to e-mail. Useful adoption is visible in changed work:

- an access request resolved without an informal referral;
- a subject access search completed without rediscovering the estate;
- a disputed figure settled through a named route;
- a change approved with known downstream consumers;
- a data-quality exception assigned to someone able to act.

The data catalogue that nobody used was rarely defeated by poor search technology. It was defeated earlier, when the organisation decided that describing the estate was the same as governing it.

The lesson of this regulatory moment is not that every organisation needs more metadata. It is that metadata must carry consequence. Until an entry changes who decides, how quickly they decide and what happens when the data fails, the catalogue remains a well-ordered account of work taking place somewhere else.