

BEYOND COMPLETENESS

The Data Catalogue That Nobody Used: What Metadata Management Keeps Getting Wrong

Why the metadata was complete, the tool was capable, and still nobody came — a reflection on the data governance push of 2018

Giovanni Leonardi

August 2018



The Data Catalogue That Nobody Used: What Metadata Management Keeps Getting Wrong

Transformation · No. 220

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

"Completeness is a supply-side measure; usefulness is a demand-side one, and in most organisations the two are almost entirely unrelated."

1. The launch that goes quiet

Every data catalogue programme I have watched reach go-live has the same first morning. The room is booked, the large screen shows the clean search bar, and someone types in a word that matters — customer, active account, net revenue. Definitions appear, each with an owner's name, a lineage diagram, a quality score. The steering group nods. The sponsor says the sentence that released the budget in the first place: at last, a single source of truth. The programme is declared delivered.

The usage report from six weeks later is where the real story is written, and it reads the same way just as often. In one representative case the catalogue held a little over four thousand documented assets and reported eighty-two per cent "completeness." In the same month it had nineteen unique users, and more than half of those were the governance team inspecting their own entries. The search logs showed hundreds of queries that returned a result and no click — someone looked, found the definition, and went to ask a colleague anyway. The catalogue was finished. It was also, in every sense that mattered, empty.

We had built the thing and mistaken building it for the goal. A populated catalogue and a used catalogue are not the same object, and the distance between them is the whole subject of this piece. Completeness is a supply-side measure; usefulness is a demand-side one, and in most organisations the two are almost entirely unrelated.

It is worth being honest about why so many of these programmes existed at all this year. The catalogue boom was not, for most boards, a sudden conversion to the value of well-managed data. It was May's regulation. With the General Data Protection Regulation now in force, organisations needed to say where personal data lived, what they did with it, and on what lawful basis — the records of processing, the ability to answer a subject access request inside a month, the promise to erase on request. A catalogue was the natural place to keep the answers. But a compliance artefact and a working tool have different finish lines: one is done when the auditor is satisfied, the other is never done at all. Funding the first while expecting the second is the original error, and almost everyone

made it.

2. Why documentation never created demand

The catalogue was scoped the way we scope anything with a budget and a steering committee: as a deliverable with a completion date and a percentage bar climbing towards it. The implicit theory of change was disarmingly simple — if we document everything, people will consult it. It is the same theory that fills shared drives with process manuals nobody opens and intranets with policies nobody reads. Documentation has never, on its own, created demand for itself.

A reference gets used when it answers a question faster than the alternatives. For most of the people we were trying to serve, the alternative was very fast indeed: turn to the analyst two desks away who has kept their own quiet list of which table is the real one, which field is trustworthy, and which report to never cite in front of the board. That informal knowledge was undocumented, unversioned, and yet completely reliable, because it was maintained by someone with a direct stake in being right. The catalogue did not lose to a better catalogue. It lost to a colleague, a spreadsheet, and a habit — and it lost on friction, not on accuracy.

3. Three ways a catalogue dies

When we look closely at the catalogues that went unused, the same three mechanisms recur, and none of them is really about the software.

- The moment of need is somewhere else. People reach for metadata at a precise instant — the second before they put a number in front of an executive, or the moment a figure looks wrong and they need to know whether to trust it. That instant happens inside the reporting tool, the query editor, the board pack. If consulting the catalogue means stopping, switching to another system, and searching afresh, the friction is higher than simply asking the person who already knows. We built a destination when what was needed was an annotation at the point of use.
- The steward had neither time nor authority. We named data stewards with some ceremony and then handed them the work side-of-desk. One steward I recall was accountable for roughly three hundred data elements on top of a full finance role; the definitions were current at launch and visibly stale within a quarter, because keeping them current was nobody's actual job description. Metadata rots faster than most people expect — an organisation restructures, a source system changes, a definition drifts — and a catalogue of confidently wrong entries is worse than no catalogue at all, because it spends the trust it was built to earn.
- The definitions belonged to no one. The hardest cases were never technical. When two functions disagreed on what an active customer was — and they always did — the catalogue faithfully recorded both definitions, or neither, or whichever owner argued hardest in the workshop. It surfaced the disagreement without any means to resolve it. A tool that exposes an unresolved argument, and then leaves it unresolved, does not build confidence; it teaches people that the official source is no more

settled than the corridor conversation, and cheaper to ignore.

A catalogue documents decisions about data; it cannot make them. Where an organisation has not decided who owns a definition, the catalogue simply publishes the ambiguity – now in a searchable, official-looking form.

4. The strongest objection, and why it is only half right

The most serious defence of these programmes runs like this: the tooling was immature, and the failures above are really failures of manual effort. The newer platforms crawl the data estate automatically, profile columns, infer relationships between tables, and suggest definitions using machine learning. Populate the catalogue by machine and the steward bottleneck disappears; adoption will follow once coverage is effortless and current.

This deserves to be taken seriously, because the first half of it is true. Automated discovery genuinely does dissolve the population problem – the very thing that made manual cataloguing collapse under its own weight. If the only failure were that stewards could not keep up, automation would be a real answer, and the pace of improvement in these crawlers over the last two years has been striking.

But it answers the wrong question. Automation solves supply, not demand. A crawler can tell you that a column is probably an email address; it cannot tell you whether this customer table is the one the finance director will defend in an audit. It can infer that two fields are related; it cannot confer authority on a definition or accountability for keeping it true. Left to run on its own, machine population does something quietly perverse: it turns a small catalogue nobody consults into a very large catalogue nobody consults, and hides the emptiness behind an impressive coverage figure. The scarce resource was never metadata. It was attention and accountability, and neither of those is something a crawler produces.

“Automation can populate a catalogue overnight; it cannot make a single person want to open it.”

5. What a living catalogue actually requires

None of this is an argument against catalogues. It is an argument against the artefact model of them. The catalogues that survive their own launch tend to rest on four commitments, and all four are organisational rather than technical:

1. Start from questions, not assets. Catalogue the twenty definitions people actually argue about before the four thousand tables that merely exist. Demand-led scope is smaller, far less impressive on a completion chart, and much more likely to be

- opened a second time.
2. Embed at the point of use. Metadata belongs where the number is consumed — beside the field in the report, inside the query tool, on the face of the board pack — not in a separate system the user has to remember to visit.
 3. Give ownership teeth. A definition owner needs the authority to settle a dispute and the protected time to maintain the entry. Stewardship offered as a side-of-desk favour is a promise made to no one.
 4. Measure use, not completeness. The only honest question is whether anyone consulted the catalogue before making a decision. A programme that reports rising coverage while usage stays flat is measuring its own activity and calling it value.

Catalogue as artefact	Catalogue as habit
Goal: a completed inventory	Goal: a consulted reference
Scope set by what exists	Scope set by what is disputed
Ownership assigned as a title	Ownership backed by authority and time
Success measured by completeness	Success measured by use before decisions
Ends as a data swamp with a search bar	Survives as the place arguments get settled

6. The mirror, not the substitute

The regulation of this year gave these programmes their money and their deadline, and in doing so handed many of them the wrong finish line — evidence for an auditor rather than a tool for a Tuesday afternoon. The catalogues that genuinely took hold were rarely the most complete. They were the ones that answered a question someone was already asking, at the moment they were asking it, with a definition that somebody was willing to own in public.

That is the harder truth beneath the unused-tool complaint. A data catalogue is a mirror of an organisation's data culture — its clarity about ownership, its willingness to settle definitions rather than let them multiply, its habits at the point of decision. Where that culture already exists in some form, the catalogue documents it and lets it scale beyond the few people who carried it in their heads. Where it does not, no degree of completeness will conjure it into being, and the catalogue becomes one more well-built system that nobody uses. The tool was never the thing that was missing.