

BEYOND THE INVENTORY

The Data Catalogue That Nobody Opened

How the year's most-funded governance tool became its most-ignored one — and why that silence is the finding, not the footnote

Giovanni Leonardi

October 2018



The Data Catalogue That Nobody Opened

Transformation · No. 222

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

“We built a card catalogue, pointed at the shelves, and called it a library.”

The system everyone approved and no one opened

The catalogue went live in the spring, not long after the new regulation took effect, and for a few weeks it was the most admired system in the building. There were screenshots in the steering pack. There was a launch note from the sponsor. Forty thousand data assets, automatically discovered and indexed, searchable from one screen — the single source of truth the organisation had been promising itself for years.

By the autumn the usage dashboard told a quieter story. On a good week, thirty people signed in, and most of them sat in the team that had built the thing. The analysts it was supposed to serve had gone back to the spreadsheet of table names they emailed around, and to the one colleague who always seemed to know which figure was the real one.

This is not one organisation's misfortune. It is the pattern of the year. Two waves of demand broke over the same tool at the same time: the regulation's insistence on a record of processing activities gave every board a reason to fund a catalogue, and the long-running dream of self-service analytics gave every data team a reason to want one. The money arrived, the platform was bought, the scanners were pointed at the estate — and the result is sitting idle in a remarkable number of places at once.

The instinct to call it an adoption problem is wrong

The reflex, when a system goes unused, is to reach for adoption: more training, a launch campaign, a line in someone's objectives that says thou shalt use the catalogue. That reading is comforting because it locates the failure in the users. It is also wrong, and acting on it wastes another year.

A catalogue populated by automated discovery answers exactly one question well: what data do we hold? That is the compliance question, and it is why the tool was fundable in a post-regulation climate. But it is not the question an analyst has at four o'clock on a Tuesday. Their question is which of these three customer tables is the one people actually trust, why do the totals differ, and who do I ask when it looks wrong? The catalogue holds table names, column types and row counts. It is silent on meaning, on trust, and on provenance-in-practice — the very things the question is made of.

A catalogue records the structure of the estate and omits its meaning. Structure can be scanned by a machine overnight. Meaning has to be curated by people, and curation is the expensive part everyone quietly deferred.

The numbers make the gap concrete. One programme I have in mind discovered forty thousand assets and curated trustworthy definitions for roughly four hundred of them — one percent — before the initiative's attention moved on to the next milestone. The catalogue was, in a technical sense, complete. In a working sense it was almost empty, and the people it was built for could feel the difference immediately even if the steering pack could not.

Metadata is a flow, not a project

The fair objection is that a catalogue is a foundation: stand it up now, enrich it with meaning later. It sounds reasonable and it fails on contact with how organisations actually behave, for two reasons.

The first is that "later" has no owner. Curation — deciding what a field means, whether it can be relied upon, who is accountable for it — is slow, unglamorous, and easy to postpone against a delivery deadline. A project that treats the build as the deliverable declares victory at go-live and disbands the team that would have done the enrichment.

The second is that metadata decays. A definition curated in March is quietly wrong by September, because a pipeline changed, a source was retired, a field was repurposed — and nobody told the catalogue, because the catalogue was a destination rather than a step in the way changes are made. A catalogue that is not wired into the flow of change does not stay a foundation. It becomes a museum of how the data used to look, which is worse than no catalogue at all, because it is confidently out of date.

There is a regulatory sting in this. The same rules that made the catalogue fundable also mistaught its purpose. A record of processing built to be shown to an auditor, and consulted by no one in between, is compliance theatre — it satisfies the letter of the obligation while delivering none of the capability the obligation was gesturing at. We congratulate ourselves on the artefact and miss the point of it.

The test that actually matters

So the question to put to any catalogue this year is not the one on the programme dashboard. It is not how many assets have we indexed? or what is our coverage? It is far simpler and far more uncomfortable: when someone had a real question this week, did they open it?

If the honest answer is no, then more assets and more training will not move it, because neither addresses why the tool is silent on the things people actually need. The failure was designed in at the start, the moment we decided that discovering the data was the

hard part and describing it could wait.

We built a card catalogue, pointed at the shelves, and called it a library. A library is the curation — the judgement about what is worth reading and what can be trusted — and that is the work we have not yet agreed to pay for.