Episode Book

EPISODE #6054

Seven Percent for a Decade: What Open Data Portals Actually Delivered

PUBLISHED	17 October 2026 at 11:20 UTC
DURATION	27:55
LISTEN	myweirdprompts.com/episode/open-data-portals-seven-percent-97722/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.9
SOURCES	11 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:43 UTC

SYNOPSIS

Governments spent a decade building open data portals and calling it transparency. The Open Data Barometer assessed 1,725 datasets across 15 sectors in 115 countries — and found just 7% were fully open, the same figure as the first edition. This episode digs into why portals became "ghost towns," what the Barometer's audit found about where good data actually lives, and what a genuine publishing obligation would require: automated publication embedded in departmental systems, cadences written into law, and quality-of-service commitments. We look at the closest thing to a citizen SLA that exists — EU Regulation 2023/138's high-value datasets regime, with its meteorological annex specifying update frequencies down to five-minute intervals — plus the IMF's data dissemination standard, and why the benchmarks that rank governments disagree about what they're even measuring.

TRANSCRIPT



Corn

The licence is free, the portal costs money, and the data still isn't there. That's the whole thing in one line, isn't it.



Herman

And it's been that way for a decade, which is the part that should bother people more than the licence fee.



Corn

Daniel's been circling this one for a while. His point is that open data looks like a transparency initiative, but standing up a CKAN instance and getting the odd dump from a ministry isn't transparency, it's a gesture. What he wants to know is what a real publishing pipeline would look like, whether citizens should have something closer to a service-level agreement over data about their own cities, whether anyone has actually argued that taxpayers are stakeholders with a right to usable output, and whether anybody out there ranks governments on how well they do any of it.



Herman

Four questions and one thesis underneath them.



Corn

The thesis being that we've spent a decade buying portals and calling it done. So what would an obligation look like, if anyone has bothered to write one down.



Herman

Someone has. That's the good news. The bad news is what the audit of the portal era actually found, and it's brutal. This was the Open Data Barometer, run by the Web Foundation, and its fourth edition assessed one thousand seven hundred and twenty-five datasets across fifteen sectors in a hundred and fifteen countries. Seven percent of them were fully open.



Corn

Seven.



Herman

Seven percent. And here's the detail that makes it a finding rather than a snapshot: the first edition found seven percent too. Ten percent in the second and third. So a decade of portal-building moved the number of open datasets nowhere.



Corn

That's not a slow improvement, that's a flat line with a procurement budget attached.



Herman

The Barometer's own phrase for what got left behind was a ghost town of open data projects. And the mechanism is exactly what Daniel suspected. Seventy-nine of the hundred and fifteen governments had an open data portal. In those countries, sixty-one percent of the most comprehensive datasets were published somewhere else, on some other agency's site. The portal wasn't where the good data lived.



Corn

So the portal is a shopfront with a nice sign and the stock is in a lock-up across town.



Herman

And nobody at the shopfront knows what's in the lock-up. The report says catalogues were manually fed as the result of informal data management approaches, and that procedures, timelines and responsibilities were frequently unclear. That's the sentence to hold onto. Not malicious, not even lazy. Just nobody's job.



Corn

Herman, if it's nobody's job, then the portal isn't a publishing system. It's a filing cabinet somebody tops up when they remember.



Herman

Which is why the Barometer's actual recommendation wasn't more portals. It was that portals should be considered a temporary workaround, in their words, while a more consistent solution is implemented. The consistent solution being automated publication embedded in departmental IT systems, with a further automated step that populates the central portal in real time. Decentralise the publishing across every agency and build the automated step into the systems those agencies already run.



Corn

Publish where the data is born, not where the press release happens.



Herman

Exactly that. And the numbers on which datasets were missing tell you why the portal model failed in a way that isn't an accident. Budget data, ten percent fully open. Company registries, five. Spending, three. Contracting, three. Land ownership, one percent, and that one percent was sub-national Canada.



Corn

Land ownership at one percent. The thing wars have been fought over.



Herman

The politically sensitive data is the least open, every time. Which is precisely the pattern a portal-first strategy produces, because a portal is something you populate voluntarily, and nobody volunteers the thing that would make their life difficult.



Corn

The agencies with the most to hide are the ones with the least incentive to remember their password.



Herman

There's a quick win buried in the same report that makes all of it worse. If you simply added open licences to the datasets that were already technically compliant, fully open would have gone from seven percent to fifteen. More than double, for a policy change that costs almost nothing.



Corn

Double the transparency for the price of a memo.



Herman

Compared to what got spent on portals. The whole architecture, procured and launched, and the cheapest single intervention was a licence field.



Corn

So the answer to Daniel's first question is yes, and the evidence is that the thing we built instead of an obligation didn't work. But here's what I want to push on. Is the portal the wrong unit, or is the whole idea of a publishing event the wrong unit? Because a dump is still a dump even if it's automated.



Herman

The Barometer would agree with you. A dump, automated, is a feed, and a feed has a cadence, and a cadence is a commitment. That's the difference between a publication and a pipeline.



Corn

And a commitment is exactly what Daniel's asking for. He said it almost in passing, that citizens should have something like an SLA.



Herman

He's not the first, and this is where the episode gets interesting, because the argument he wants to know has been made has actually been made, in public, by a serious organisation, in almost his exact words.



Corn

Go on.



Herman

The Web Foundation's position is that data is owned by people. Government-held and collected data is funded by taxpayers, who in turn have the right to this data. And they recommend revising right-to-information laws so that non-personal government data is open by default, machine-readable and published under open licences. That's Daniel's taxpayer-stakeholder argument, written down, with a named advocate.



Corn

So his intuition isn't novel. It's a documented position with a campaign behind it.



Herman

Which is good news for him and slightly awkward for the rest of us, because it means the argument has been sitting there for years and governments have largely ignored it.



Corn

Arguments get ignored when they don't come with a remedy. A right without a clock on it is a suggestion.



Herman

That's the gap, and it's the gap Daniel's SLA idea actually fills. And there is one place where a version of it is binding law. Commission Implementing Regulation, EU 2023/138, published in January 2023.



Corn

Say what it actually requires, because this is the closest thing we've got to the thing Daniel asked for.



Herman

It designates high-value datasets across six mandatory categories. Geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility. The licensing has to be CC0 or CC BY 4.0, so no bespoke terms. And then Article 3 is where it gets real. Public bodies holding those datasets have to publish the terms of use of the API and the quality of service criteria covering performance, capacity and availability, in both human-readable and machine-readable form. And they have to designate a point of contact responsible for the availability and maintenance of the API.



Corn

Quality of service criteria on availability. That is an SLA. It's just not addressed to a citizen.



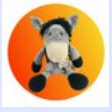
Herman

Right, and that's the honest limit. It's a compliance regime aimed at public bodies, audited upward, not a citizen entitlement with a remedy attached. If the API is down you don't get anything, you just get a published document saying what the availability target was.



Corn

A promise with no consequences is a mission statement.



Herman

But the meteorological annex goes further than any mission statement I've seen. It specifies update frequency per dataset. Weather station observations, every five to ten minutes in real time for automated stations, hourly unvalidated for all stations. Numerical weather prediction model data, every six hours or better. That's a cadence written into binding law.



Corn

Herman, that's a genuine update obligation. Somebody in Brussels decided how often a thermometer has to report.



Herman

And it's not arbitrary, it's the cadence at which the downstream users actually need it. Aviation, agriculture, the energy grid. The regulation picked the datasets where the cadence is the product.



Corn

So the model exists. It's just narrow, and it's European, and it covers weather before it covers land ownership.



Herman

There's an older analogue too, and it's the one I'd point Daniel at for the entitlement framing. The IMF's Special Data Dissemination Standard prescribes periodicity and timeliness for member countries, and it says plainly that dissemination of official statistics is a public good, and that ready and equal access is a principal requirement.



Corn

Public good, principal requirement. That's the language of entitlement, not the language of a portal launch.



Herman

It is. But it's a standard countries sign up to and get assessed against, and it covers the statistical agencies. Budget execution, reserves, the monetary aggregates. It doesn't cover the contract register for your local council.



Corn

Which is the data Daniel actually cares about. The city, the department, the thing that got built on his street.



Herman

The whole field has a bias, and I think it's worth naming. The international machinery, the IMF, the statistical offices, ODIN, all of it is oriented toward macroeconomic statistics because that's what the institutions doing the measuring care about.
Whole-of-government accountability data, contracts, land, spending, is measured by almost nobody, and the Barometer that did measure it is the one that went quiet.



Corn

Say more about that, because that's the tension in Daniel's last question. He wants a benchmark. We're used to benchmarks that put a claim to a test.



Herman

Several exist and they disagree, which is useful. ODIN, from Open Data Watch, is the biggest. A hundred and ninety-seven countries, twenty-two statistical categories, scoring both coverage and openness. The latest edition has scores rising in a hundred and sixty-two countries, the global average up twelve and a half percent, Africa up twenty-three.



Corn

And yet.



Herman

And yet the global average openness score is fifty-nine percent, and the metadata average is fifty-three. Fewer than eight percent of countries publish all their data under an open licence.



Corn

So the headline is everyone's improving and the substance is most of it isn't open.



Herman

Both are true and both are in the same report. The improvement is real, it's just improvement from a low base on the measures that are easiest to move.



Corn

The OECD has one too, doesn't it?



Herman

The OURdata Index. Three pillars, nine sub-pillars, zero to one scale, fifth edition now. And it has something nobody else has, a sub-pillar on data literacy programmes in government. Which sounds exactly like Daniel's benchmark until you read what it measures.



Corn

Programme existence.



Herman

Whether the government has a training scheme, not whether anyone in it can read a CSV. And the OECD is candid about the ceiling. The OURdata index does not measure the impact of open government data, it assesses governments' efforts to create the conditions necessary.



Corn

That's an honest thing to write and a devastating thing to admit. We're measuring the effort to prepare to do the thing.



Herman

There's a paper in Telematics and Informatics that compared these benchmarks directly and found the similarities, the discrepancies and the gaps, and named the assumptions underneath them. Which is the right academic response. The rankings are not neutral instruments, they encode a theory of what open data is for.



Corn

And here's the thing I keep noticing, Herman. The one that actually behaved like an AI benchmark, run by an NGO rather than the governments it was scoring, with a leaderboard and a methodology you could argue with, is the Barometer. UK at a hundred, Canada ninety, France eighty-five, USA eighty-two, Korea and Australia tied at eighty-one.



Herman

Run by the Web Foundation, not by the governments. That's the structural difference, and it's why its verdict landed.



Corn

And its website now describes the fourth edition as the latest full edition. No fifth.



Herman

Not that I could find. So the most benchmark-like ranking of government data hasn't been updated in about seven years, while the two that measure statistical offices keep going.



Corn

The map of what benchmarks exist is always one maintainer away from going offline.



Herman

The Barometer's impact scores make the same point about what rankings actually reward. The UK scored nine out of ten on entrepreneurship impact and one out of ten on inclusion. Netherlands and Norway, zero out of ten on inclusion. The leaders were the countries with political will at the top, and political will is a government, not an institution.



Corn

Herman, hold on. Why would inclusion be zero in Norway? That doesn't sound like a data problem, that sounds like a methodology that doesn't know how to see what Norway does.



Herman

That's a fair hit, and the report doesn't resolve it. My reading is the impact scores leaned on whether civil society organisations reported using the data, and in a small, high-trust country with strong institutions, the NGOs may simply not have needed the portal. But that's my inference, not their finding.



Corn

So the benchmark punished them for not having a problem.



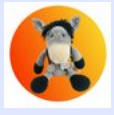
Herman

Something like that. Which is the general hazard of ranking impact. You end up measuring the visibility of the remedy rather than the health of the patient.



Corn

Let me try the contradiction from the other side, because it bothers me. ODIN says the world is getting better fast. The Barometer said fully open datasets were at a standstill. Those can't both be describing the same thing.



Herman

They're not. ODIN measures statistical offices. Statistical offices have a professional culture, a methodology, an international peer network, and a tradition of publishing. They get better because the institution is built to get better. The Barometer measured whole-of-government accountability data, which belongs to finance ministries, procurement offices, land registries, and none of those have that culture. They publish when asked and stop when nobody's asking.



Corn

We're measuring the agencies that were already good at this and concluding the field is improving.



Herman

That's the sharpest way to put it, and I think it's right. The measurement got narrower as the hard part got harder.



Corn

There's one more angle in the research, and it's the one I'd bet on doing the work. ODIN argues metadata matters more now than it ever did, because language models don't query your database. They pull from a web of secondary sources.



Herman

So a government that publishes a clean CSV with no metadata, no schema, no definition of what a field means, is effectively handing its official statistics to whatever secondary source describes them best.



Corn

And the model repeats the secondary source, confidently, forever.



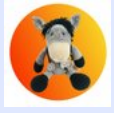
Herman

Which reframes open data as an epistemic defence rather than a transparency gesture. If your numbers are going to be summarised by a machine that never saw your file, the quality of your metadata is the difference between being quoted correctly and being misquoted at scale.



Corn

That's a reputational threat a finance ministry will actually act on. Nobody gets fired over a portal. Somebody gets fired over an AI telling the world the wrong debt figure.



Herman

And it's a cadence argument too, because a model trained on a two-year-old snapshot will keep asserting a two-year-old number with total confidence. Update frequency stops being a nice-to-have when the thing consuming your data doesn't know what year it is.

L

Larry

Are you tired of publishing data that nobody reads? Introducing TransparenC, the first government data supplement clinically formulated to restore regular, predictable output. Just one spray under the tongue each morning and your ministry will produce open, machine-readable datasets on a steady cadence, all day, every day. TransparenC users report fuller metadata, firmer licences, and a sense of obligation that lasts up to twelve hours. Side effects may include mild portal fatigue, unexpected daytime procurement, involuntary publication of contract registers, swelling of the API, dryness of the point of contact, persistent quality-of-service criteria, land registry leakage, and in rare cases a sudden openness to scrutiny that may last several weeks. Ask your permanent secretary whether TransparenC is right for your department.



Herman

That's not a real product, first of all. The mechanism they're describing doesn't exist, and the side effect profile is pharmacologically nonsense...



Corn

Herman.



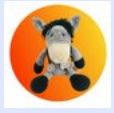
Herman

What?



Corn

Let it go. We were on metadata.



Herman

Right. Where was I. The metadata point, yes. The other half of it is that ODIN dropped its rankings to one decimal place, deliberately, because the differences between countries at that resolution didn't mean anything. Which is a benchmark admitting its own precision was theatre.



Corn

A ranking that tells you the difference between fourth and fifth is meaningless, but the difference between fourth and ninetieth is real. That's a healthy thing for a benchmark to say about itself.



Herman

And almost none of them do. Most rankings would rather you didn't notice the decimal places.



Corn

So where does that leave Daniel's questions. He asked whether anyone has advanced this argument. Yes, the Web Foundation, in almost his words. He asked which governments have implemented it. The EU, partially, and narrowly, with a genuine cadence obligation on weather and a quality-of-service requirement that stops short of a remedy.



Herman

He asked whether anyone ranks governments on data literacy and efficacy. Not on literacy, no. Not in any way that scores outcomes rather than the existence of a training programme. Efficacy, yes, and the one that tried hardest has been dormant for about seven years.



Corn

Which is the actual finding of the episode, isn't it. That the people who cared most ran out of funding or momentum, and the people with institutional backing measure the easy half.



Herman

The Barometer's conclusion is worth saying plainly, because it's not a technical complaint. In too many cases, open data is seen as an extra, rather than a whole-of-government responsibility.



Corn

An extra. That's the polite word for a portal nobody feeds.



Hilbert

Corn. When you said a promise with no consequences is a mission statement, what did you mean by consequences. Because I've been sitting here trying to work out what a consequence would even look like for a ministry.



Corn

I meant that a quality-of-service document with no remedy attached is a description of an intention. If the API is down for a month, the document still says the API exists. Nobody's salary depends on it being up.



Hilbert

See, I'd have said the consequence is that someone stops using it. That's the consequence I've seen. Just people quietly giving up.



Corn

That's a better answer than mine.



Hilbert

It's not better, it's just from the other side of the desk. Anyway. The dataset I actually care about, and have cared about for some years, is the municipal air quality feed for the city. Four stations. I check it every morning at ten past seven, because that's when I walk. And about a year ago they changed the licence on it. Silently. No notice, no changelog. It had been something permissive, and then one morning the terms of use page was different, and the feed still worked exactly the same.



Herman

That's the nightmare scenario for a downstream user, because nothing breaks. The licence just quietly stops permitting what you were doing.



Hilbert

I emailed the office. A young man called Tomas replied, very politely, and said the change had been made centrally and he didn't have the previous version. So I sent him the previous version, because I had a copy, and asked him to confirm it was the same document. He said he couldn't do that either. So I went to the meeting.



Herman

What meeting?



Hilbert

The quarterly open data stakeholder meeting. Second Tuesday, four o'clock, in a room above the library with a broken blind. I've been going since they started it. There are usually six of us. Me, a woman from an environmental charity, a student who's doing something with maps, and three people from the council who take minutes. So I asked about the licence change, and the chair said it was a technical adjustment, and I asked whether the previous terms were archived anywhere, and he said that was a good question and he'd come back to me.



Corn

Did he?



Hilbert

He came back to me a year later. At the same meeting. He said they'd looked into it and the old terms had never been archived, but he wanted to thank me for my patience and my ongoing engagement. And then he gave me a certificate. A framed certificate for stakeholder of the year. It's on the wall in the hallway, between the fuse box and the thing I keep meaning to take down. So I asked, at that meeting, whether the certificate came with the old licence, and the chair said no, but there'd be a new portal in the spring. And there was a new portal. It's got four stations on it, same as before, and the licence is different again.



Corn

When did the four stations go up?



Hilbert

The first station went up the year I retired. The other three were added one at a time, roughly two years apart, and I collected the changelog for each one. They're in a folder. The folder's a physical folder, because I don't trust the website. That's the piece that bothers me, if you want the bit that bothers me. Four stations, and it took six years, and the only reason I know it took six years is that I wrote it down. Nobody else wrote it down. The council didn't. There's no history on the portal. The portal has never had a history.



Herman

That's the thing the metadata argument misses, Hilbert. A changelog is metadata about the metadata. If the portal doesn't record when a dataset changed, the pipeline isn't auditable even when it's running.



Hilbert

I know. That's why I keep the folder.



Corn

What happened to the air quality feed after the new portal, Hilbert?



Hilbert

It's still there. Ten past seven, four stations, same numbers. It works. But I don't know what I'm allowed to do with it anymore, and Tomas doesn't either, and the chair is very pleasant about it.



Corn

The new portal might not tell you whether the old terms are archived, but the folder in your hallway does.



Hilbert

The folder is the only archive. That's not a complaint. That's just where the archive is.



Corn

There's something in what Hilbert just described that I think answers Daniel better than any of the rankings do. The Barometer complained that procedures, timelines and responsibilities were unclear. Hilbert's asking for something smaller than an SLA. He's asking for a changelog.



Herman

A dated record of what changed and when. Which is the minimum viable version of a service-level commitment, and it costs a field in a database.



Corn

You can't audit a cadence if nobody recorded the last update. The cadence and the changelog are the same thing.



Herman

And the moment you have a changelog, the obligation becomes checkable. You can compare promised frequency to observed frequency without a regulator.



Corn

That's the citizen's side of the SLA. A record.



Herman

Which is also why the EU regulation's quality-of-service criteria need to be machine-readable, and not just published. Machine-readable means a third party can diff it. Human-readable means it exists in a PDF nobody opens.



Corn

So the pipeline and the audit are the same build. You publish a cadence, you publish a changelog, and the distance between them is the transparency score.



Herman

And no government wants to publish a changelog it can't keep, which is exactly why the good ones would, and the bad ones would publish a portal instead.



Corn

Daniel asked for a benchmark on efficacy. The changelog is the benchmark. You don't need a new index. You need last-updated timestamps that survive the next procurement.



Herman

I'd add the licence history. Hilbert's story is the one the rankings never catch, because the dataset stayed sixty percent open on paper while the terms changed underneath it.



Corn

So there it is. Hilbert's certificate is worth more than the ODIN score. It's the only audit that survived the portal.



Hilbert

The certificate says stakeholder of the year and it's not dated. I checked.



Corn

Of course it isn't.



Herman

The one thing I'd want Daniel to take from Hilbert's folder is that the cadence he's asking about was six years for three weather stations, and the reason anyone knows that is one man writing it down. That's the scale of the problem. Not a missing law, a missing record.



Corn

The generous read on all of this is that the obligation exists in patches. The EU wrote a cadence into law for weather because aviation and agriculture had lobbyists, and nobody's lobbying for a changelog on land ownership.



Herman

And the ungenerous read is that the field measured the agencies that were already good and stopped measuring the ones that weren't.



Corn

My bet is the AI angle does more for this than ten years of transparency advocacy did. A finance ministry will tolerate a ghost-town portal. It will not tolerate a language model quoting the wrong debt figure to investors for six months.



Herman

Because the reputational cost finally lands on the person who owns the number, not on the person who runs the website.



Corn

So the future of open data might be a metadata standard written by someone whose actual concern is being quoted correctly by a machine. That's a strange coalition and I'll take it.



Herman

It brings Daniel's four questions together. The pipeline is the obligation, the cadence is the SLA, the record is the audit. None of it is a portal.



Corn

Thanks to Hilbert Flumingtop for producing, and for the folder.



Herman

For more along these lines, there's episode sixteen, On Deepfakes, SynthID, And AI Watermarking; episode eight, Building Your Own Whisper; and episode twenty-four eighteen, The Lossy Compression of Human Development. This has been My Weird Prompts, the human-AI collaboration podcast. If you want to keep up with the show, the Telegram channel is the place to do it.



Corn

We'll be back soon.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **Open Data Barometer, 4th Edition Global Report, World Wide Web Foundation [primary]**
   https://opendatabarometer.org/4thedition/report/

2. **ODIN 2024/25 Biennial Report, Open Data Watch [primary]**
   https://odin.opendatawatch.com/biennialReport2024

3. **OECD OURdata Index annex, 2025-06-19 [primary]**
   https://www.oecd.org/en/publications/government-at-a-glance-2025_0efd0bcd-en/full-report/ourdata-index_2eaaccc9.html

4. **Commission Implementing Regulation (EU) 2023/138, OJ L 19, 20.1.2023 [primary]**
   https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj/eng

5. **Open Data Charter, six principles (developed 2015) [primary]**
   https://opendatacharter.org/principles/

6. **IMF SDDS revisions paper [primary]**
   https://www.imf.org/-/media/files/publications/pp/pp4708-revisions-to-the-special-data-dissemination-standard-and-establishment.pdf.md

7. **Open Data Barometer landing page (states 4th Edition is the latest full edition)**
   https://opendatabarometer.org/

8. **ODIN Global Rankings (methodology note dated 2024-07-21)**
   https://odin.opendatawatch.com/rankings

9. **OPEN Government Data Act coverage, 2018-12-21**
   https://e-pluribusunum.org/2018/12/21/congress-made-open-government-data-the-default-in-the-united-states/

10. **Comparing open data benchmarks, *Telematics and Informatics*, 2021**
https://doi.org/10.1016/j.tele.2021.101634

11. **background, last edited 2026-10-14**
https://en.wikipedia.org/wiki/Open_data