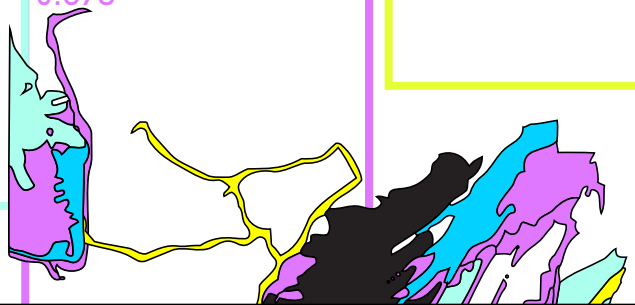
Open Knowledge

THE TECH WE WANT | AI LEARNING LABS

Connecting AI to Government Data

A field guide by the Open Knowledge Foundation on how to increase accuracy, improve provenance and build trust in AI responses



SEPTEMBER 2026

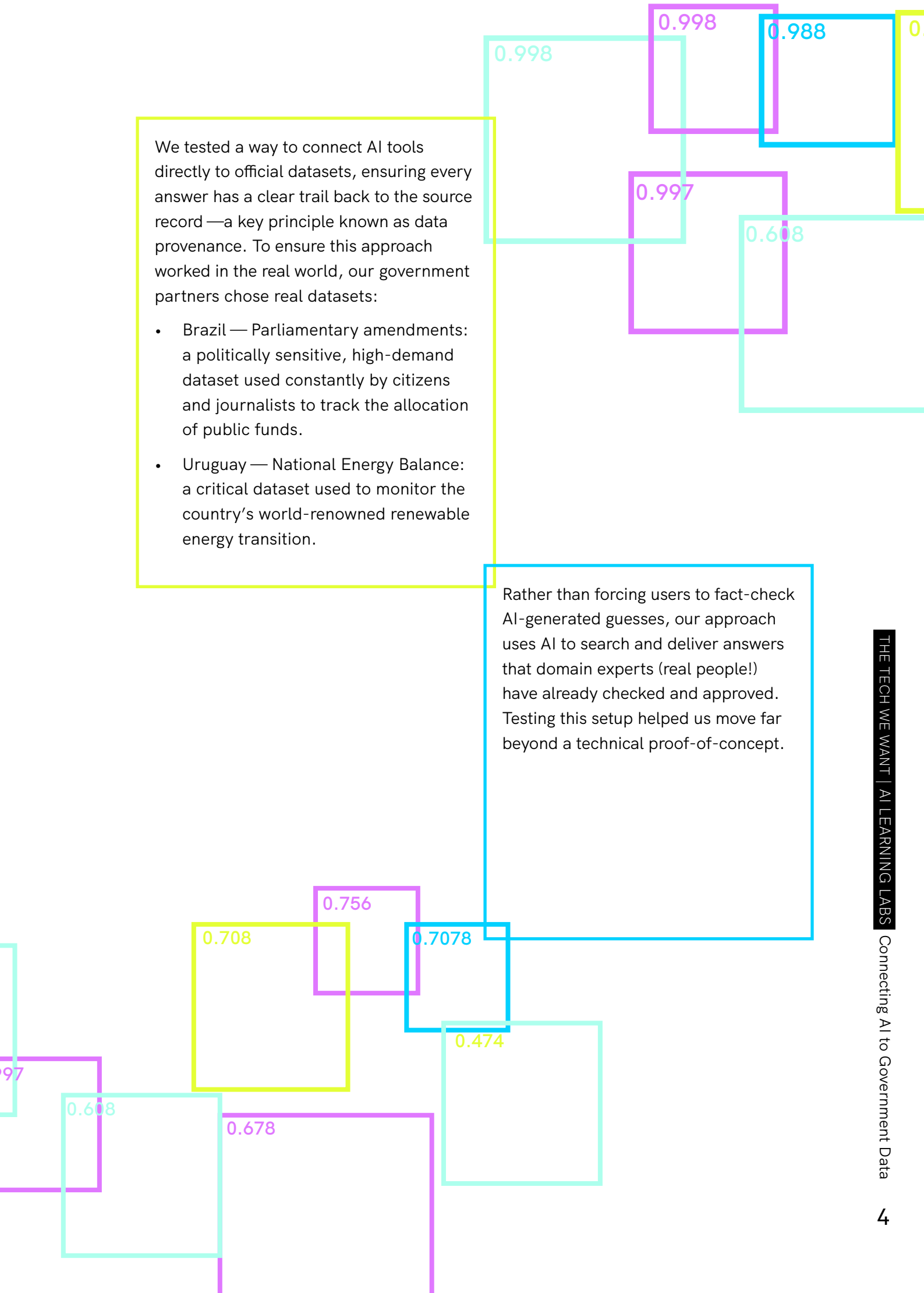
INDEX

ReadMe

Available data ≠ Accessible data

Public data is essential for accountability, participation, and informed decision-making. Yet, citizens, researchers, and journalists too often must navigate complex portals, understand the data structure, master technical queries, and invest significant time simply to extract answers.

Over the course of four months in 2026, the Open Knowledge Foundation's AI Learning Labs tackled this challenge by connecting open-source AI to public open data portals. The goal was to enable citizens to ask natural-language questions and 'talk' to national data. The challenge was to create a process that minimised the risk of false information (hallucinations) and allowed responses to be traceable.



We tested a way to connect AI tools directly to official datasets, ensuring every answer has a clear trail back to the source record —a key principle known as data provenance. To ensure this approach worked in the real world, our government partners chose real datasets:

- Brazil — Parliamentary amendments: a politically sensitive, high-demand dataset used constantly by citizens and journalists to track the allocation of public funds.
- Uruguay — National Energy Balance: a critical dataset used to monitor the country's world-renowned renewable energy transition.

Rather than forcing users to fact-check AI-generated guesses, our approach uses AI to search and deliver answers that domain experts (real people!) have already checked and approved. Testing this setup helped us move far beyond a technical proof-of-concept.

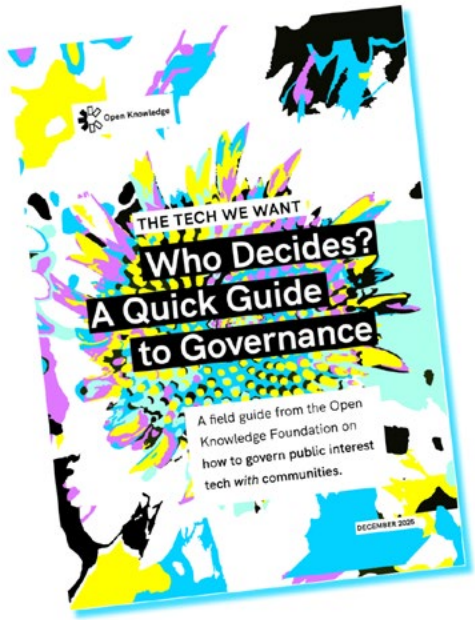
See the previous guides
in this series…

Read This Before You Build — A field guide on how to build public interest tech *with* communities

'We need to rethink how we build and govern digital tools that truly work for *us* to ensure that we have options, now and in the future.'

ReadMe | On learning and unlearning

Who Decides? — A field guide on how to govern public interest tech *with* communities



'Community governance is what enables a project to evolve through the visions and needs of more than just the people who write the code.'

ReadMe | Why governance matters

Our ‘Recipe’

Team + Infrastructure + Timeline

To help you replicate it in your context, here is what you need for a successful pilot based on our AI Learning Labs experience.

1. Team

(Minimum viable setup)

- **Technical developer** (to connect the AI to the data)
- **Domain expert** (to clarify dataset definitions, assist with data analysis and flag errors)
- **Project coordinator** (to coordinate testing and keep the project moving)
- **Government partners** (with a dedicated team to manage collaborative scoping, decisions, and approvals)

2. Data & Infrastructure

(You do not need to build the system from scratch. Consult our detailed Technical Developer Guide.)

- Accessible dataset: An official database, portal, or structured API containing clean public records.
- Complete metadata: Data dictionaries, glossaries, and background notes explaining column titles, measurement units, and known data gaps.

3. Timeline & Process

(For a lean team reusing existing infrastructure)

- **[Weeks 1-2] Setup:** Align on goals, choose simple initial datasets, and set up the connection between the AI and the database.
- **[Weeks 3-5] Testing & Iteration:** Testers ask questions, domain experts review answers for accuracy, and developers update rules and glossaries in iterative loops.
- **[Week 6] Reflection:** Pause to review team feedback, document what worked, and plan improvements before scaling to complex data.

Success

Recommendations for humans and machines

FOR HUMANS

HUMAN UNDERSTANDING OF THE DATA IS KEY

Connecting the AI directly to official datasets successfully allowed it to pull information directly from the source. But for answers to be correct, the system also needs context: what terms mean, what units values use, which time periods figures cover, etc. It takes human experts who understand the data to determine what conclusions the data actually supports.

What we recommend:

Develop the technical foundation alongside the people who understand the data and keep them involved.

TESTING NEEDS THE RIGHT PEOPLE

Checking AI-generated answers against public data is not a simple pass/fail exercise. It takes time and requires people who understand the dataset well enough to spot answers that sound plausible but are not actually supported by the data. The mix of testers matters: government insiders were more likely to catch wrong figures, while external civil society testers noticed when answers were confusing or assumed too much background knowledge.

What we recommend:

- Involve both domain experts and external users in testing.
- Budget enough time for human review, including time to trace unclear or questionable answers back to the source.

TRACEABILITY ALONE DOES NOT GUARANTEE CORRECTNESS

While the pilots confirmed where the answer came from, they also showed that a source link does not prove the full answer is correct. In one Uruguay example, the system referred to “climate factors” when explaining changes in energy data. Across 84 test questions, testers marked 32 answers as correct, while 21 were marked as partially correct or uncertain. The most visible recurring issue was a coverage or retrieval gap: the system could not provide an answer even though testers believed the information existed in the data.

What we recommend:

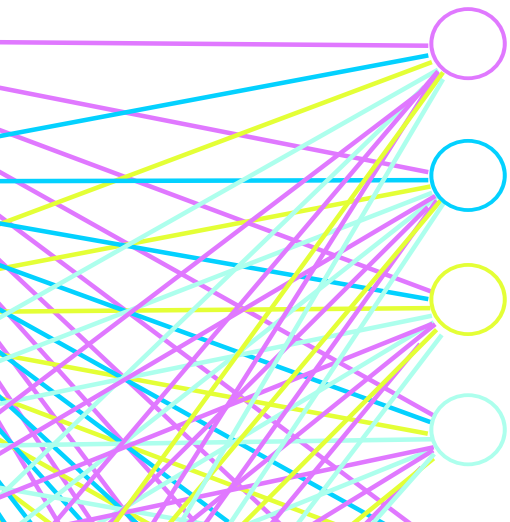
- Keep traceability at the centre of the design and test whether the source actually supports the full answer, rather than just verifying that a source link appears.

REAL-LIFE QUESTIONS DRIVE SYSTEM IMPROVEMENTS

Pilot partners helped provide questions that people actually ask about the datasets. Those real-life questions improved the system: they showed where glossaries were missing terms, where everyday language didn’t match official data terms, and where the assistant needed strict limits so it would decline to answer instead of guessing. For example, “What is an amendment?” showed the need to include core concepts in the glossary. A question about “emendas pix” (a reference to a popular payment system in Brazil) showed the need for a dictionary that maps informal terms to dataset terminology.

What we recommend:

- Build the glossary and terminology dictionary earlier, and make them part of the system context. Watch out for informal language (words, abbreviations and acronyms).
- Include edge cases and high-risk questions, especially where a wrong answer could mislead users.



FOR MACHINES

SEPARATE FACTS FROM INTERPRETATION

Several testers noticed that the assistant sometimes mixed information from the datasets with its own general training data. When this happens, data provenance is lost—users can no longer tell where government records end and the AI's own assumptions begin. Updating system prompts helped the model separate data from interpretation, but this alone only reduces the problem; ongoing validation remains necessary.

What we recommend:

- Require the assistant to distinguish between data-backed facts and interpretation and limit the AI to answering only from the information given.
- Make that distinction visible in the answer itself.

ARITHMETIC CALCULATIONS NEED EXTRA VALIDATION

The pilots showed a clear difference between two kinds of questions. Questions asking the system to find an existing value (e.g., "What was the value in this year?") were much more reliable. Questions asking the system to calculate a new value (e.g., percentages, shares, or year-on-year changes) were considerably riskier. In the pilots, this was one area where testers reported answers that were numerically wrong—and because wrong calculations are presented convincingly, they are easy to miss.

What we recommend:

- Make sure important percentages, shares, and changes are calculated in a controlled way before they are shown to users.
- Do not rely on AI models to generate database queries or perform math on the fly, as language models frequently make arithmetic errors when calculating on demand.

Error

Limits that we could not cross

The “blank chat” problem

Because an empty chat box doesn’t tell users what filters apply, what categories exist, or what years are covered in the data, testers often didn’t know what questions they were “allowed” to ask. Furthermore, when the system couldn’t answer a prompt, users couldn’t easily tell why: was the data missing, did the AI misunderstand, or was the question simply outside the system’s scope?

Potential workarounds:

- Explicitly surface the system’s scope, available datasets, and limitations before the user starts typing.
- Guide the user’s queries by providing visible examples of what can and cannot be answered.
- Test with real users from civil society and government to ensure system boundaries are easily understood by non-technical audiences.

Conversation flow needs more systematic testing

We tested multi-question conversations, but have not yet analysed conversation flow in enough detail to draw strong conclusions across both pilots. Answer quality can vary within a conversation, including accurate answers followed by inaccuracies or “No answer” results. But to understand why this happens, future evaluation should look more closely at the content of each exchange: whether the user asked a follow-up, changed topic, used different terminology, challenged the answer, or asked something the data could not support.

Potential workarounds:

Evaluate conversation flow as its own test area, rather than merely treating it as a sequence of separate questions.

Pathways

6 steps to a data provenance framework for AI agents

1. Make data and metadata usable by AI

For AI systems to use public data responsibly, they need more than a file or API endpoint. Public data needs clear metadata, shared definitions, documented limitations, and traceable sources.

2. Show the source

The system itself must guarantee that outputs are grounded strictly in source data. If an accurate answer cannot be retrieved, the system should decline to answer rather than guess.

3. Train the AI to recognise jargon

Ensure your system includes domain dictionaries or semantic mappings that translate everyday language (informal terms, shorthand, and industry slang) into official data categories using administrative taxonomy.

4. Start small to build trust

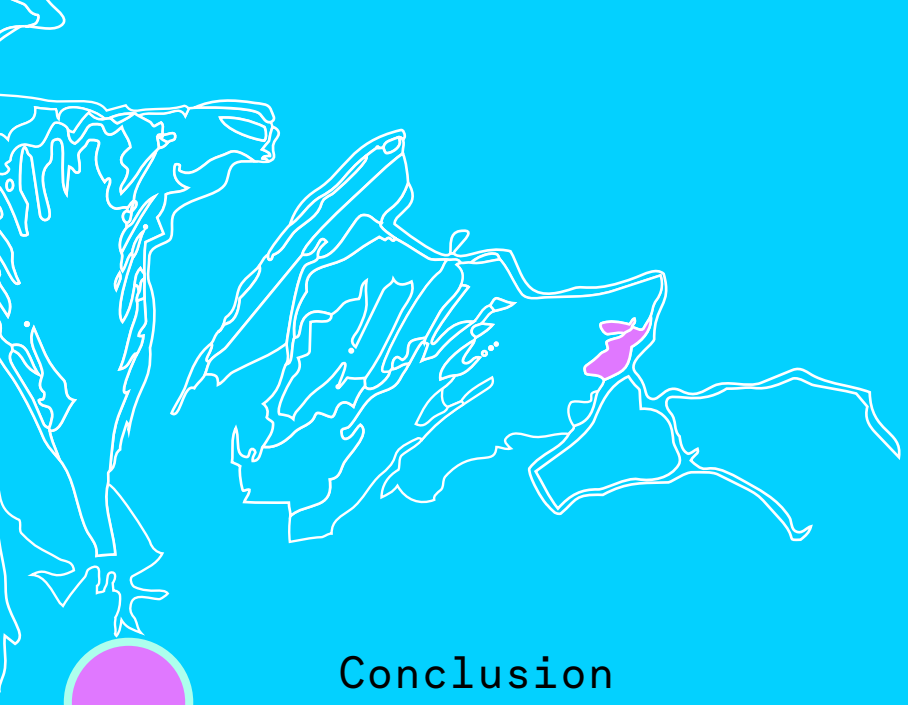
Do not attempt to make your most complex dataset AI-ready on day one. A working pipeline on a simple dataset builds institutional trust and creates a blueprint you can later scale to more difficult domains.

5. Budget for meaning

Make glossaries, metadata, and dataset limits part of the system. Plan time for definitions, informal terminology, units and time periods, known data gaps and questions the data cannot answer.

6. Keep humans in the loop

Domain experts and government data teams are needed throughout the entire design and development cycle. Civil society members, journalists, and non-technical readers also bring crucial perspectives to make the source trail clear.



Conclusion

Should open data communities adopt AI?

The definition of Open Data includes machine-readability. Today, that statement should be expanded to **AI readiness**.

Our pilots integrating traceable AI into government information suggest that open data portals should NOT be replaced by chatbots. Rather, they suggest that the tools we use to publish and query open data must evolve to **match how people and systems ask questions now**.

For data providers and open data teams, this means **thinking beyond publication documentation** for people to find online. For civil society, researchers, and other users, it means developing solutions for the AI answer to show the source, and also **the source to actually support the answer**.

Open Knowledge
FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

By: Open Knowledge Foundation

Thank you to the team who worked on this project: Andrés Vázquez, Kannika Thaimai, Lucas Pretti, Patricio del Boca
A special thanks to the teams from the Brazilian Office of the Comptroller General (CGU) and the Agency for the Electronic Government and for an Information Society in Uruguay (AGESIC)

Illustrations and design: Constanza Figueroa Bustos

September 2026

Recommended citation style:
Open Knowledge Foundation (2026): Connecting AI to Government Data - A field guide on how to increase accuracy, improve provenance and build trust in AI responses

Open Knowledge Foundation (OKFN) is a not-for-profit organisation incorporated in England & Wales with company number 05133759. okfn.org

Openness is our foundation.

Patrick J McGovern FOUNDATION

We are thankful for the support of the Patrick J. McGovern Foundation. Learn more about its funding programmes here.

#TheTechWeWant

A collective conversation about practical ways to build software that is useful, simple, long-lasting and focused on solving people's real problems.
Join us:
okfn.org/the-tech-we-want

#AILearningLabs

Catalysing learning and developing replicable methods to help organisations build AI skills, use AI responsibly, and develop their own AI projects. Learn more: **okfn.org/ai-learning-labs**

Get involved

Run a pilot: Interested in testing this? Reach out to us at **info@okfn.org**

Join the conversation: Have lessons or questions from your own work? Connect with other practitioners in our **Open Knowledge Forum.**