

## **Open Data Does Not Mean Accessible Data: The Next Frontier of Transparency in the Public Sector**

When people talk about innovation, the conversation often goes straight to technology. In the public sector, however, innovation is about who delivers greater value to citizens. Brazil's Access to Information Law is a good example, as it established public disclosure as the rule and confidentiality as the exception. More than simply responding to requests, the law introduced the proactive disclosure of information of public interest, regardless of whether anyone requested it, strengthening active transparency.

Openness expanded further with the Open Data Policy and the plans through which each government agency began defining what it would publish and when. More recently, the volume of data available on the federal portal has grown significantly, and Brazil reached 8th place in the OECD Open Data Index.

This trajectory has also involved building capabilities within the State itself. The National School of Public Administration, for example, offers several training programs in data science and artificial intelligence, including initiatives specifically aimed at women. These efforts matter because the teams within government agencies are the ones who understand the databases, organize the information, and create the conditions for data to be used in new ways.

These are real achievements. But one question remains: who can actually use this data?

### **The Remaining Barrier**

I work with public health data and encounter this problem frequently. Information may be available, but that does not necessarily mean it is easy to access. Answering something that appears simple may require downloading large files, understanding unintuitive codes, combining datasets that do not communicate with one another, and, often, knowing how to program.

For those who work with this data every day, the path becomes familiar. For those outside this environment, it does not. A citizen, a health council member, or a manager who does not work with data may find the information available and still be unable to turn it into an answer.

It is no coincidence that, in the OECD index, Brazil's lowest-performing pillar is support for data reuse. Publishing data has advanced faster than making that data effectively usable.

This is where a new challenge for transparency emerges. Opening data remains essential, but it may no longer be enough. The distance between published data and citizens' questions needs to be reduced.

What if people could simply ask?

This is precisely where a new generation of tools can help.

Today, it is already possible to transform a question written in everyday language into a query against a database. Instead of searching for files, figuring out which table to use, or writing code, a person can simply ask: "How much did my city spend on medications in recent years?" The system interprets the question, queries the data, and presents the result.

This is already beginning to happen in Brazil's public sector. In 2026, Brazil's National Treasury launched Siconfi.IA, which allows users to ask questions in natural language about the accounting and fiscal data of the federal government, states, and municipalities, while also helping users understand concepts and regulations.

It may appear to be merely a new way of querying a database, but the change is greater than that: it reduces the barrier between those who have a question and those who can reach an answer.

The tools needed to move forward already exist. Data science makes it possible to organize large volumes of information, connect datasets, and create new forms of querying. More recently, artificial intelligence has added another possibility: enabling interaction in everyday language, without requiring users to understand the structure of the data or know how to program.

This does not mean that all government agencies are at the same stage. For some, the challenge is still to better prepare their data for new forms of use. Others are already ready to move toward tools that allow citizens to interact directly with data.

### What Needs to Happen Behind the Scenes

The simpler the experience becomes for the person asking a question, the more important the work behind the answer becomes.

A public database may contain hundreds of fields. And a seemingly simple word, such as "value," can mean different things: estimated value, contracted amount, or amount actually paid.

Recent studies on these tools show that one of the main challenges is identifying which information truly answers the question. It is not enough to query the data; it is necessary to understand what the data means.

For this reason, making queries easier requires documenting the meaning of information, recording rules and changes over time, and bringing together technology professionals and those who understand the data in their daily work. This knowledge usually exists within the teams themselves, among those who know why two apparently identical numbers do not match or which rule changed during a given period. Turning this knowledge into documentation is also innovation.

And the answer must be auditable. Anyone who receives a number should know where it came from and when the data was last updated. Showing the path behind an answer is what separates reliable information from an unsupported number.

### Not Everything Needs to Be Answered

Making queries and data combinations easier also requires caution. Not every dataset that can be combined should be combined. Information that does not identify anyone when viewed separately may reveal an individual when combined with other information, especially in areas such as healthcare or in small municipalities.

A good tool is not one that answers every question, but one that recognizes when there is enough information to provide a safe answer and allows users to verify where that answer came from.

### Closing Thoughts

The Access to Information Law changed what the State makes public. The open data policy expanded the amount of information available. The next step is to change what citizens are able to ask and understand.

I envision a future in which different public databases can be consulted in an integrated way and anyone can ask questions using everyday language. A future in which someone asks, “How much did my city invest in healthcare last year?” and receives a clear answer, with the data source, the update date, and the ability to verify how that result was obtained.

Reaching this point depends not only on technology. It will require combining tools, organizing data, involving teams, and establishing clear rules so that answers are useful, reliable, and secure.

Open data that only specialists can access is transparency only halfway achieved. Enabling anyone to turn a question into information they can understand and verify is, in my view, the next frontier of transparency.

#### Further Reading

Shi, L., Tang, Z., Zhang, N., Zhang, X., & Yang, Z. (2025). A Survey on Employing Large Language Models for Text-to-SQL Tasks. *ACM Computing Surveys*, 58(2), Article 54, 1–37. <https://doi.org/10.1145/3737873>.

Ali, M., Maratsi, M. I., Lachana, Z., Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2026). Talk to Open Data: Enabling User Interaction with Open Government Data Using LLMs, RAG and Smart Agent Technologies. In P. Rupino da Cunha & M. Themistocleous (Eds.), *Information Systems. EMCIS 2025. Lecture Notes in Business Information Processing*, vol. 573 (pp. 15–31). Springer. [https://doi.org/10.1007/978-3-032-18487-0\\_2](https://doi.org/10.1007/978-3-032-18487-0_2).