



Why the World Needs SDMX and StatGPT

**SMDX GLOBAL CONFERENCE
ROME, MARCH 2025**

Jeff Danforth, IMF

Outline

- Motivation
- Out-of-the box tools – cautionary tale
- 3 steps to follow so statistical agencies use AI to help users find their data.
- Next steps

Motivation

- Many statistical organizations underperform in data quality areas of accessibility and serviceability.
- Weaknesses include:
 - Description of data in technical terms
 - Dissemination in multiple places, vintages, and formats
 - Official source of information often lost
- AI tools, like StatGPT, accessing official statistics modelled using SDMX can drastically improve this.

Out of the box tools

- Tools like ChatGPT can bridge the gap between statistical language and natural language but caution warranted

Can you produce a table of economic growth for the g7. Can you take the data from the IMF's WEO?

 Certainly, here is a table summarizing the real GDP growth projections for the G7 countries, based on the International Monetary Fund's (IMF) World Economic Outlook (WEO) as of October 2024:

Country	2023 Growth (%)	2024 Growth (%)	2025 Growth (%)
United States	2.1	2.7	1.9 [2.2]
Canada	1.3	1.6	1.8 [2.4]
United Kingdom	0.5	1.1	1.5 [1.1]
Germany	0.2	0.0	0.7 [0.8]
France	1.0	1.3	1.5 [1.1]
Italy	0.7	0.7	1.0 [0.8]
Japan	2.0	1.0	0.9 [1.1]

These projections indicate that the United States is expected to experience the highest growth among G7 nations in 2024, while Germany is projected to have stagnant growth during the same period.

↓

INTERNATIONAL MONETARY FUND

- Correct response 34 percent of the time when 10 prompts were repeated in the same conversation, 17% correct when the request was made using unique conversations.
- Significant variation across countries: United States responses correct 93% of the time and Germany never had a correct response.
- Responses are plausible looking making them hard to dismiss outright.

Out of the box tools

- New versions play it much safer regarding data questions

Can you generate a table of economic growth rates for G7 countries taking the data from the latest IMF WEO Report.

I currently don't have direct access to the internet or live databases like the IMF WEO Report, so I can't pull the latest data directly. However, I can guide you through how to access it and provide a general framework for the table you're asking for.

Steps to access the IMF WEO data:

1. **Visit the IMF WEO Page:** Go to the [IMF World Economic Outlook \(WEO\) page](#).
2. **Select the Latest Report:** Choose the most recent report from the list.
3. **Download the Data:** Typically, the IMF provides both a PDF summary and an Excel file with the full data. The Excel file will have detailed data on GDP growth for all countries.
4. **Filter for G7 countries:** The G7 countries are:
 - Canada
 - France
 - Germany
 - Italy
 - Japan
 - United Kingdom
 - United States

Once you have the data, you can either share the relevant part of the dataset with me, or I can help you structure the table based on typical GDP growth rate entries.

If you have the data handy, feel free to upload it, and I can help organize it for you!

- New version of ChatGPT no longer generate tables with incorrect data.
- Users are provided the steps to access the data or asked to provide it to the chat bot so it can organize it.
- User still does not get the accurate data they were looking for.

How can statistical agencies use AI to help users find their data?

1. Make data AI ready using SDMX

Data modelled using SDMX ready to be consumed by AI agents. Benefits include:

- Easily understood (by AI) structures
- Well-identified with descriptive metadata and semantic relationships
- Semantic versioning provides transparency and predictability for updates
- Discoverable and machine accessible via SDMX API endpoints in multiple formats
- SDMX community committed to advancing the standard and providing training
- Backed by statistical agencies with subject matter expertise for their data

SDMX can be further extended to improve AI compatibility (TWG is on it!):

- Data harmonization within and between organizations
- Model Context Protocol (MCP) making AI agents available to any front-end

Modelling your data using SDMX and making it available via a SMDX API endpoint is not enough....

2. Place data in an index of certified AI-ready Statistics

Introducing the **Global Trusted Data Commons**

- The GDC is global index of official, machine-readable, and metadata-rich statistics from trusted sources.
- It acts as the underlying "truth layer" for GenAI applications to draw from when responding to a request for official statistics.
- Envisioned to be the global index of official statistics that have been made AI “ready” by official statistics producers.

2. Place data in an index of certified AI-ready Statistics

Introducing the **Global Trusted Data Commons**

- Readiness centers around **consistency and completeness**:
 - Data modeling using SDMX (2.1 or 3.0, accessed via SDMX1 library)
 - Structural metadata that is unambiguous, clear and descriptive
 - Structural artifacts with consistent naming conventions, descriptions that indicate business purpose
 - Open access via public APIs with low latency, service-to-service authentication if needed

3. Use a Gen AI application (like StatGPT) to find and retrieve data

- StatGPT is an open-source generative AI interface.
- Purpose built to query the Global Trusted Data Commons.
- Serves as **demonstrative tool** showing how GenAI can be used to responsibly interact with official statistics
- Development steered by the SDMX community and national statistical agencies allowing for custom UI/UX and guardrails.

3. Use a Gen AI application (like StatGPT) to find and retrieve data

“Can you generate a table of economic growth for G7 countries using the IMF WEO”

Can you generate generate a table of economic growth rates for G7 countries taking the data from the latest IMF WEO Report.



> View Processing Steps

Below is the table of economic growth rates (annual percent change in real GDP) for G7 countries, based on the latest available data from the IMF World Economic Outlook (WEO) report (last updated March 2025):

Country	2020	2021	2022	2023	2024	2025	2026	2027
Canada	-5.04	5.95	4.19	1.53	1.53	1.38	1.56	1.71
Germany	-4.10	3.67	1.37	-0.26	-0.23	-0.05	0.92	1.45
France	-7.59	6.83	2.62	1.12	1.07	0.64	1.02	1.19
United Kingdom	-10.30	8.58	4.84	0.40	1.10	1.08	1.41	1.54
Italy	-8.87	8.93	4.82	0.72	0.73	0.44	0.84	0.57
Japan	-4.17	2.70	0.94	1.49	0.08	0.55	0.58	0.63
United States	-2.16	6.06	2.51	2.89	2.80	1.83	1.74	1.98
G7 Aggregate	-4.22	5.82	2.62	1.85	1.71	1.23	1.38	1.58

Source: IMF World Economic Outlook (WEO), last updated March 2025

URL: [IMF WEO Dataset](#)

- Correct response 100 percent of the time.
- Source clearly stated and linked.
- Fully integrated into traditional faceted-search data explorer.
- StatGPT demonstrates what's possible but is just one AI application in a large ecosystem.

Agent-to-agent connections (MCP) will allow ChatGPT, Google, Microsoft, Claude, etc. to benefit from the agents the statistical community controls.

What's Next

- A StatGPT advisory committee has been established:
 - Consists of SDMX sponsors and statistics agencies
 - Guides the creation of the Global Trusted Data Commons
 - Steers the development and testing of StatGPT
- In line with SDMX Community principles, **StatGPT** is open source.
- Test the latest version of the **Global Trusted Data Commons** using StatGPT
- Check out development partner's (**EPAM**) booth
- Continue to work the private sector to create synergies between tools
- Enquire about **onboarding** your data.

StatGPT GitHub Repo



<https://github.com/epam/statgpt/>

Try Official Statistics Commons
(12 datasets from IMF)



<https://portal-stg.statgpt.org/>

We need your help to make the GTDC the trusted source for official statistics!!

ANNEX

	StatGPT	Global Trusted Data Commons
What is it?	Open-source product that can perform talk-to-your-data on top of SDMX datasets	SaaS platform that operates on top of SDMX datasets of Global Trusted Data Commons member organizations (the implementation of StatGPT)
How do I use it?	Deploy the instance to your infrastructure and index your datasets	Join the GTDC
What is the data?	Your SDMX data	Data from member agencies
Who owns the instance?	You	GTDC
Can I index my private data?	Yes	No
Adjustment options	<ul style="list-style-type: none"> • Back-end • Front-end • Content 	<ul style="list-style-type: none"> • Content
Integration options	<ul style="list-style-type: none"> • Embed into your web-site • Run as standalone tool • REST API • MCP 	<ul style="list-style-type: none"> • Link to Official Statistics Commons • REST API • MCP

Criteria for entry in the Global Trusted Data Commons

SDMX

- Version 2.1 or 3.0 access via SDMX1 library
- SDMX-ML for [structure](#) endpoint format, ML or JSON for [data](#) endpoint format
- Required endpoints include code lists, concept schemas, DSDs, dataflows, available constraint, data

Connectivity

- Public API with low latency
- No authentication or service-to-service authentication available
- Ability to control rate limit

Structural metadata

- Structural metadata that is unambiguous, clear and descriptive
- Versions follow versioning conventions relevant to the associated standard
- DSDs: dimensions, attributes, and measures with consistent naming conventions
- Concept schemas: concepts, names and descriptions indicate business purpose
- Code lists: items should have unique names, hierarchies well defined with no contradictions