

“Beyond the buzz: what does ‘AI-readiness’ really mean for official statistics?”

Eric Anvar

22nd Meeting of the CSSP, 18-19 June 2025



oecd.org/data



data-explorer.oecd.org



OECD Statistics



@OECD_Stat



» Main take-aways from the survey on AI-readiness

1. AI to regenerate the age-old business case for streamlining and automating statistical process – only requiring to do it much more *systematically* and *at scale*.
2. AI to catalyse *higher efficiency* in statistical production and *better data accessibility* – with also opportunities for new products and services.
3. *Main gap* is with the data and metadata (machine readability, harmonised concepts, standard formats, availability of structural and referential, well-crafted metadata).
4. *Priority areas of investment* in 2025: Skills and Platform.

» Focus on AI-ready (meta)data

1. AI-ready data ready to be consumed by AI services and agents

Harmonised, structured, augmented with rich metadata to optimally feed local or public LLMs → *Crucial importance of capitalising on and extending statistical standards in place (especially SDMX).*

2. AI-ready data to train or fine-tune AI

Multi-source (traditional + alternative); Multi-lingual; Compliant (with legal and ethical criteria) → *Potentially augmenting training/testing data by human experts with LLM-generated, synthetic training sets.*







» What to *together* rather than *alone*?

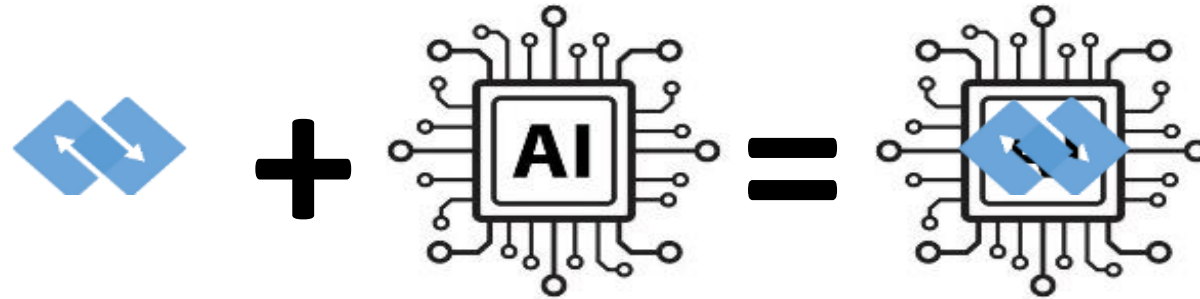
1. Facilitate **co-investment** in frontier AI-related R&D
2. Enrich **statistical standards** with AI-ready features
→ Facilitate the interoperability with **big tech players**
3. Develop the AI-readiness **frameworks**
4. Develop AI skills and foster a **culture of innovation**



» What, Where, How can OECD contribute?

What	Where	How
Facilitate co-investment in frontier AI-related R&D		Offers a space for co-investment (today development and maintenance of open source → AI prototypes, training, development and maintenance).
Enrich statistical standards with AI-ready features		Define AI-oriented features, develop guidelines and support the implementation. SDMX is already a space for technical dialogue with tech players. Broader engagement with tech players could also take place through industry fora such as AI Alliance/OTDI, and, for political discussions, in statistical fora (UNSC, CSSP...) as well as AI policy fora (GPAI, ...).
Facilitate interoperability with big tech players	?	
Develop the AI-readiness frameworks		Continuing and amplifying the excellent collaborative on AI for official statistics initiated in the UNECE HLG-MOS context.
Develop AI skills and awareness in workforce		AI capacity building program targeting low/middle income countries and statistical offices by Parsi21.

» Example: the SDMX+AI initiative



1. Launched in the wake of the 2023 SDMX Global Conference, with [SIS-CC](#) members and partners. See [inception report, March 2024](#).
2. 2 promising co-investment opportunities under review:
 - ➔ StatGPT (led by IMF, introduced at CSSP 2024, follow-ups March, May 2025)
 - ➔ AI-boosted Data Explorer (NL search, follow-ups March, May, June 2025)



offers a unique setting to coalesce funding (seed money + dev + maintenance) from multiple statistical organisations to deliver open assets

» HLG-MOS generative AI project

Generative AI and Official Statistics UNECE Workshop

12-14 May 2025, Geneva, Switzerland



From 12 to 14 May, over 100 participants from national statistical offices, international organizations, academia, and the private sector gathered in Geneva to explore the integration of generative AI into production and its governance.

Over three focused days, the workshop combined strategic discussion with technical depth, addressing emerging use cases, risks, governance models, and future pathways for collaboration, check: [Generative AI and Official Statistics Workshop 2025 | UNECE](#)

UNECE HLG-MOS offers a unique forum to develop share experience, develop guidelines and build capacity in the area of generative AI for official statistics.

The background features a series of thin, grey, wavy lines that flow from left to right. These lines are punctuated by small, dark grey dots, creating a sense of movement and data flow. The overall aesthetic is clean and modern, typical of a corporate or technical presentation.

Annex: CSSP AI-Readiness Survey results

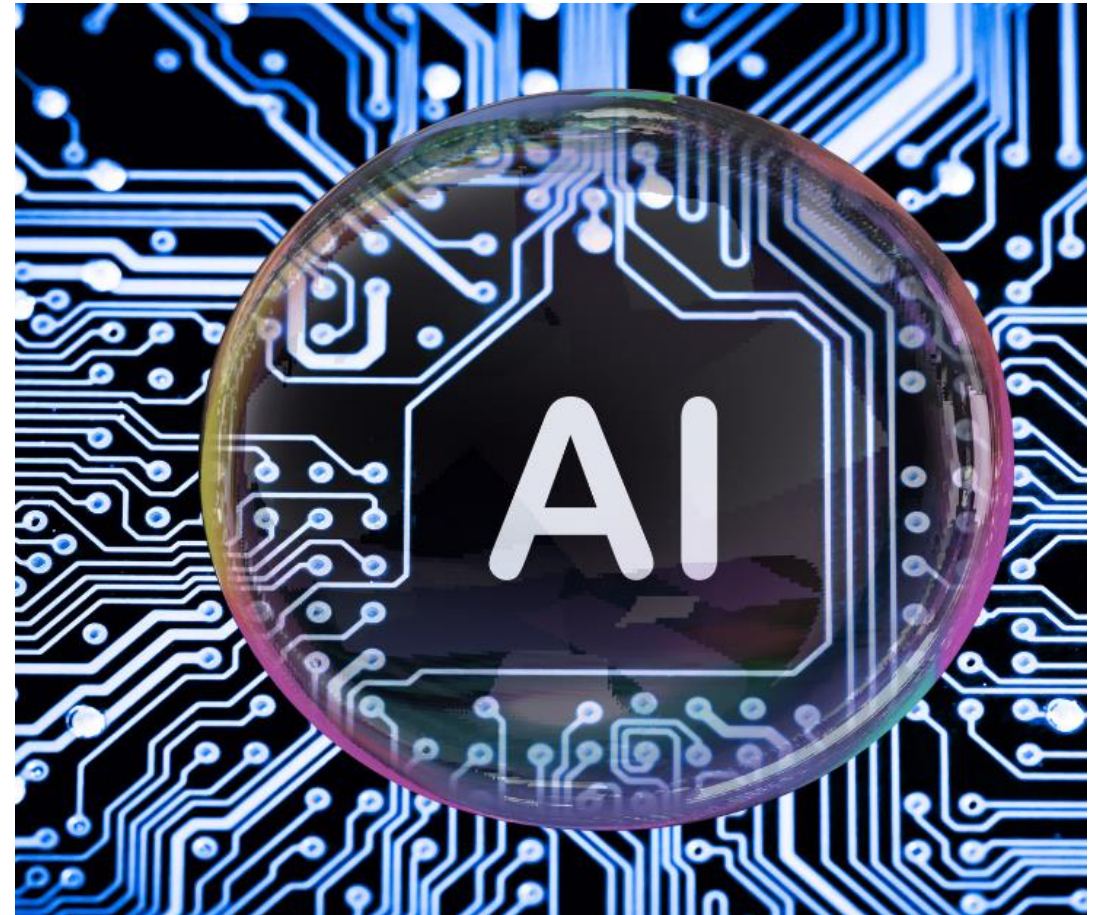
» Overview Pre-CSSP Meeting Questionnaire

Recipients (52 Invites)

- CSSP Members and Associates
- OECD Accession Candidate Countries
- OECD Key Partners, Ukraine

Responses

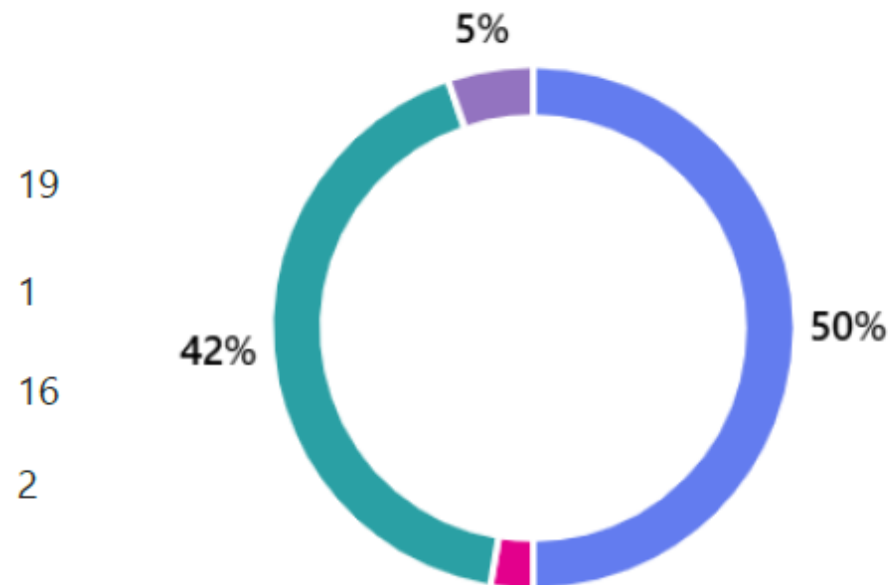
- 36 countries (incl. EU) responded
- Response rate 69%



» What is “AI-ready” data?

1. Which of the following best reflects how your organisation thinks about ‘AI-ready’ data?

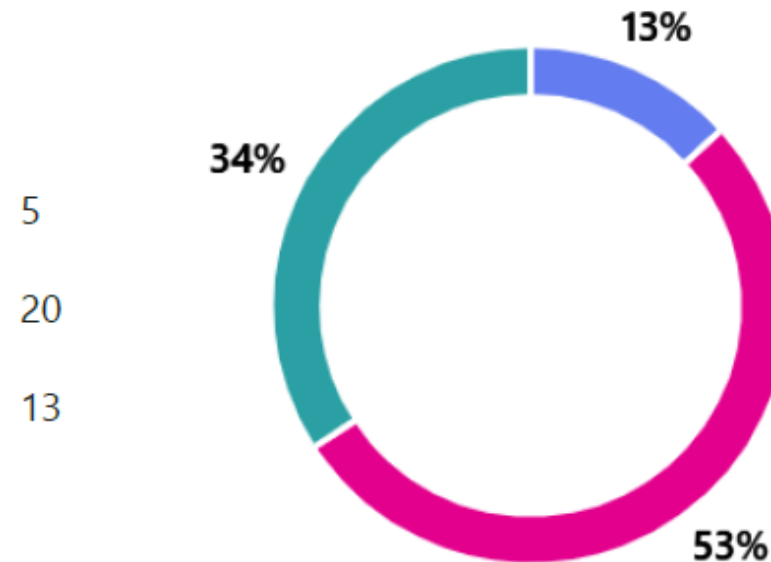
- Data optimized for AI: data that are machine-readable, structured, and accompanied by rich...
- AI-enabled data: data that result from AI-enhanced processes, such as automated classification,...
- All of the above
- Other



» A strategy to become “AI-ready”?

2. Has your organisation developed any strategy for making data ‘AI-ready’?

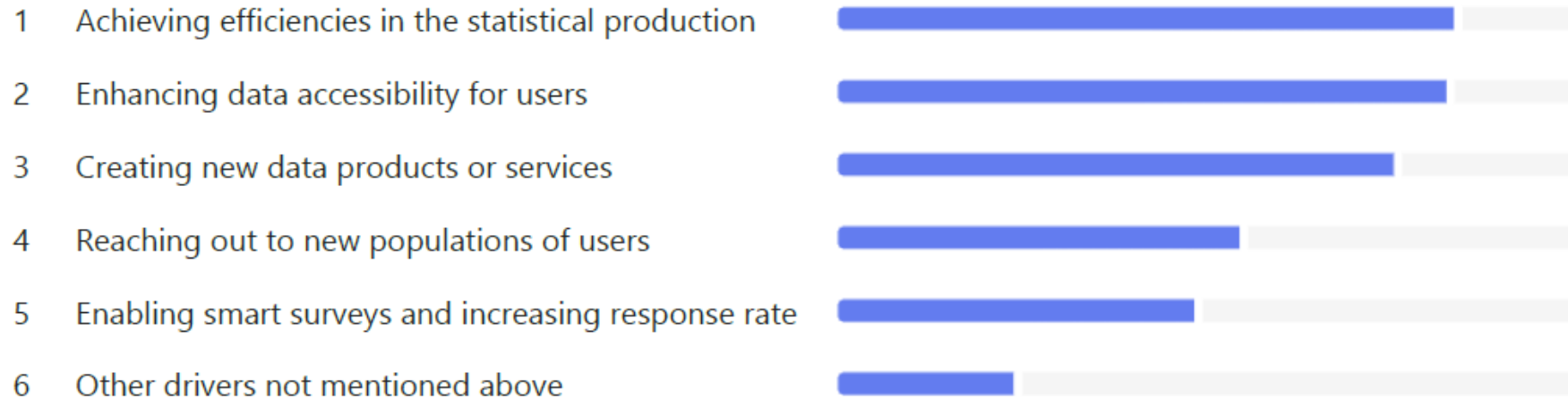
- Yes
- Nothing specific to “AI-readiness” – rather a continuation of pre-existing practice
- In progress



» Main drivers?

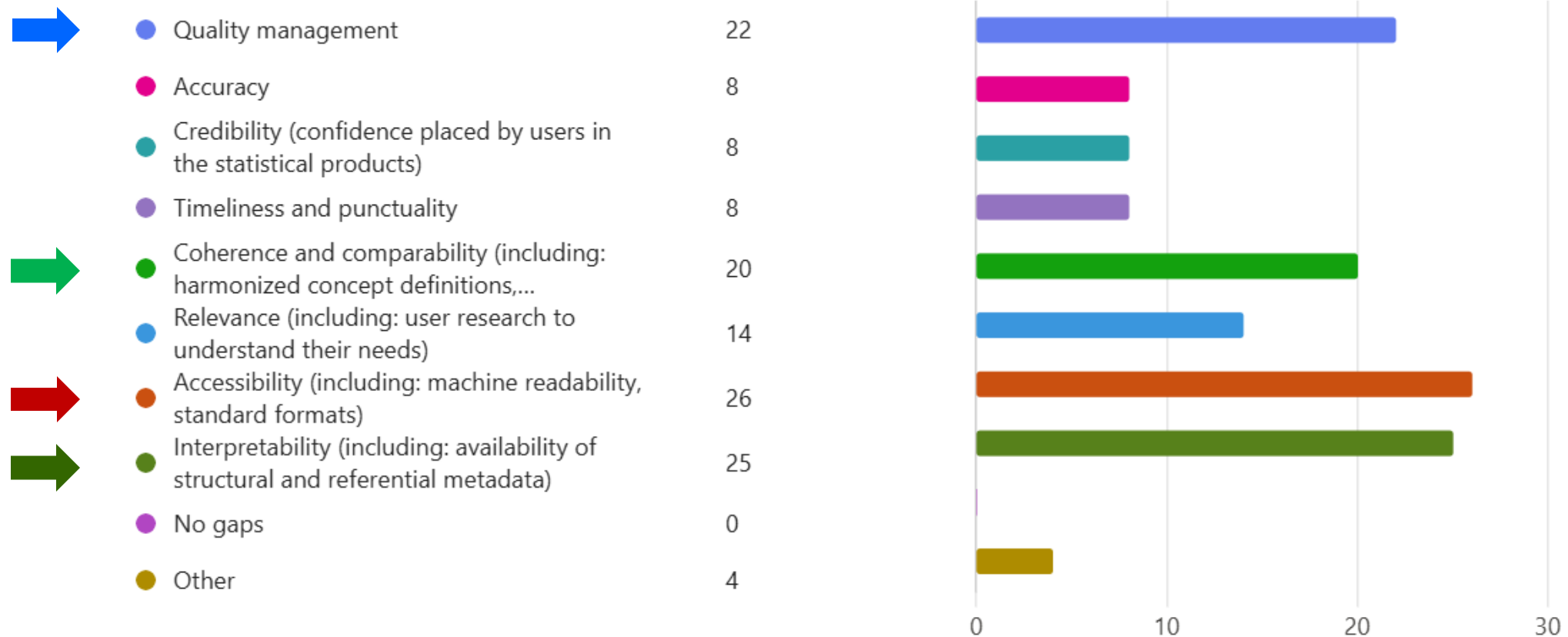


4. In making your data 'AI-readiness', what do you see as the main value drivers?



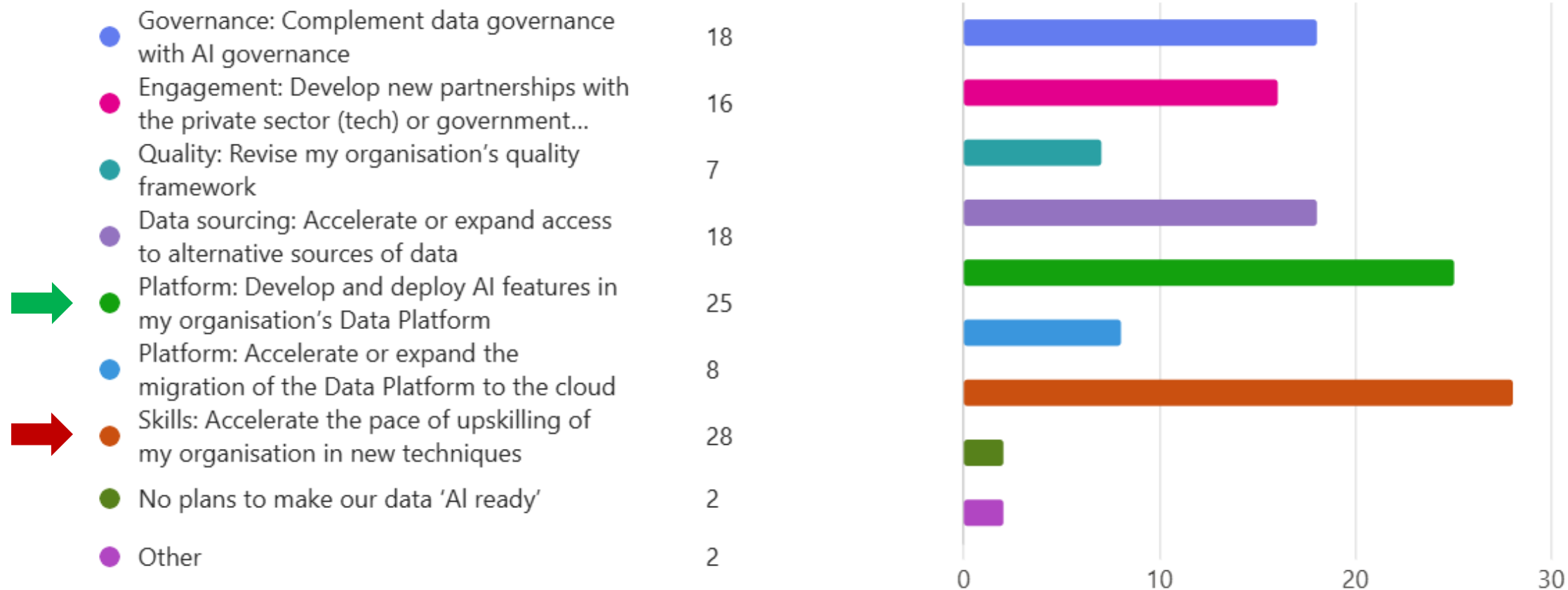
» Gaps in “AI-ready” data

5. Were are the gaps in your data ‘AI-readiness’; which dimensions of your quality framework should you prioritise?



» 2025-2026 priorities

6. In which of your core capabilities do you intend to prioritise investment in 2025-26 to make your organisation and data 'AI-ready'?



» Cited components of an AI strategy

1. **AI-ready data (consume):** Harmonised, augmented, structured, metadata-rich for optimally feeding local or public LLMs
2. **AI-ready data (train):** Multi-source (traditional + alternative); Multi-lingual; Compliant (with crucial legal and ethical criteria)
3. **AI-ready infrastructure:** Modern platform for (meta)data management and dissemination (cloud-native, interoperable). Dependent on the use case, space for commercial LLM and locally-hosted open LLM
4. **AI-ready governance:** At the cross-road of data and digital governance
5. **AI-ready workforce:** Recruit, train, empower staff to innovate with AI
6. **AI-ready administration:** Optimise tasks (recruitment...); AI procurement





oecd.org/data



data-explorer.oecd.org



OECD Statistics



OECD Podcasts



OECD_Stat

