Data Use Benchmarking Report
A Comparative Analysis

This report is the outcome of a data use benchmarking exercise run in the autumn of 2023 as part of the collaborative work by the Innovation Growth Lab (IGL) and the European Innovation Council. It aims to provide an overview of how data is utilised by leading innovation funding agencies to support innovation. An in-depth examination of the methods and approaches of these organisations serves to guide future directives in an age where the application of data for strategic insights has become critical.

October 2024

Authors:
David Ampudia Vicente
Sara García Arteagoitia
George Richardson

Summary of findings

The benchmarking exercise explored how many innovation funding agencies use data for decision-making and strategic operations. Several challenges as well as trends and best practices emerged through surveys and interviews across different departments. The agency designations were anonymised for this report and their participation in the exercise does not necessarily constitute an endorsement or validation of the conclusions.

Overarching challenges

- **Data is underutilised across the board:** Although agencies have access to extensive data, it remains underutilised due to limited capacity, mismatched skills, and low demand from policymakers. While those managing the data are often aware of these challenges, broader recognition across an agency may be lacking, particularly among those that can drive change. Non-overlapping mandates often create areas of oversight, sidelining valuable datasets that could inform strategy.

- **Ensuring data quality rather than availability is the challenge:** Data quality poses a bigger concern than data availability. Missing or erroneous data hinders accurate analysis for many agencies, though some have taken proactive steps with dedicated staff to address data accuracy concerns.

- **Agency sizes shape data capabilities and risk exposure:** Larger agencies have broader capabilities but face more complex needs. Smaller agencies benefit from direct communication but are often reliant on key individuals for expertise. This concentration of skills, coupled with inconsistent funding for key positions, leaves agencies vulnerable to disruptions.

- **Siloed operations hinder data integration:** Data integration is challenging due to fragmented operations. Silos often result from a lack of joint ownership or teams brought together without prior collaboration. Limited incentives and analytical fluency to see data's broader value also contribute to these silos.

- **Institutional communication approaches shape data projects:** The methods by which data projects are requested, processed, and communicated vary considerably among agencies. Several contend with frequent ad-hoc or short-notice requests, with only a small subset having structured processes. Many rely on key individuals in excess.

- **Limited data literacy curbs enhanced data utilisation:** Many organisations have the necessary skills, but their use depends on demand, which is tied to policymakers' data literacy. There is a need to enhance this literacy, especially concerning techniques such as causal inference and predictive modelling.

Emerging Trends & Best Practices

- **Data curation and maintenance**: Agencies are recognising the importance of well-curated data repositories. Some have established robust data management systems, while others are developing data warehousing solutions to better integrate and process isolated data streams into centralised or hybrid systems.

- **Knowledge management and ownership:** Structured conversations around knowledge management are helping agencies define clear data practices. These efforts are fostering an environment where decision-makers increasingly rely on data, improving data literacy and encouraging buy-in for more advanced methods. This, in turn, generates better evidence for decision-making.

- **Data-driven decision-making for agile operations:** Agencies are increasingly leveraging internal and external data sources to streamline processes and improve routine decision-making. A subset of agencies, some with IGL's support, are beginning to use experimental evaluations to test and refine how data insights can reduce operational redundancies and improve process efficiency.

- **Tiered data access based on expertise**: Agencies are adopting tiered data access models, granting experts deeper access to granular data while providing guided access to others via an open FAQ interface. This approach optimises data usage while maintaining data integrity and reducing request redundancy.

- **Advanced techniques for unstructured data:** Agencies are starting to explore natural language processing and test the use of large language models to analyse unstructured data, such as grant applications and reports. Some are currently commissioning this work externally, but plan to bring it in-house as they build up their data analysis capabilities, sometimes with assistance from IGL.

- **Sustainability in skill dependency**: Agencies are taking steps to address the over-reliance on key individuals by implementing upskilling programmes, knowledge transfer protocols, and process documentation. These efforts help ensure continuity and reduce the risks associated with losing critical expertise.

- **Bottom-up initiatives to drive data upskilling:** Grassroots efforts are emerging within agencies to advance data utilisation and management. Data governance and upskilling working groups demonstrate how internal motivation can drive the adoption of data-driven approaches, even without top-down directives.

Introduction

Until recently, organisations that have funded science and innovation have collected data as part of the record keeping process of managing programmes. Today, there is a recognition that such data also has high value for strategic decision-making. Agencies that have the ability to link internal data with external data and derive insights using data science methods gain a clear edge.

As data science and AI methods continue to mature, funding agencies can make the most of the insights from their data. To learn how data is used effectively, IGL and the European Innovation Council (EIC) developed a programme of work to compare and learn from different agencies through individual and collective gathering of insights. This study is, to the knowledge of both organisations, the first benchmarking study focused specifically on data use, capabilities, and strategy across multiple innovation agencies, leading to concrete discussions on improving data-driven decision-making.

IGL suggested involving external stakeholders through a benchmarking study with several innovation agencies. This process included initial surveys followed by detailed interviews to gather more in-depth feedback. The study focused on two main questions: how well an agency uses data science, and how effectively it applies this data to its strategy. This approach aimed to provide individual feedback to each agency and also highlight common challenges or areas of interest.

The following insights from the benchmarking exercise served as a starting point for action. Agencies could see where they stood compared to their peers, identifying areas for improvement and collaboration. These findings shaped the [Innovation Data Dialogues](#), a series of peer-to-peer workshops focused on key topics like data governance, policy frameworks, and experimental methods. By using the benchmarking results, agencies were able to engage in targeted discussions, share best practices, and work towards more effective and collaborative data use in innovation funding.

The first section of this report discusses the approach to benchmarking that underlies this report and the framework through which data science proficiency is understood. The second section provides a summary of the outcomes across all categories within this framework. The final section contains a more detailed breakdown of benchmarking results within those categories. The report concludes with the strategic implications of the findings and considerations for other organisations seeking to undertake a benchmarking exercise.

Methodology

The study employed a two-pronged methodology:

Written survey: A questionnaire was used to capture foundational data while minimising participant burden. A short set of questions was designed for quick, easy responses, focusing on data availability within the agency and its data analysis capabilities.

In-depth interviews: Following the questionnaire, live calls were conducted to explore the responses. These discussions used open-ended questions to gather deeper insights into each agency's data use, the challenges they face, and their future objectives.

Building on this approach, IGL developed a nine-point scoring system to assess data science proficiency across five categories. Points were allocated using a set rubric, based on questionnaire responses and interviews. The sections assessed are outlined below:

Data Skills & Capabilities	
Composition	Data team's size, mandate, and specialisation
Skills	Proficiency in foundational and advanced data skills
Capabilities	Process-oriented capabilities and supporting data infrastructure
Data Collection & Integration	
Integration	Ability to merge diverse, complementary data streams
Use	Breadth and depth of data use across functions
Quality	Data accuracy, timeliness, and quality control processes
Analytical Methods	
Basic	Use of descriptive statistics and visualisations
Intermediate	Application of standard econometric tools (ie. regression models)
Advanced	Deployment of complex analytics methods (ie. machine learning)
Use Cases	
Impact	Integration of data into strategic planning and decision-making
Efficiency	Role of data in optimising processes and tracking metrics
Insight	Use of data to anticipate trends and understand ecosystems
Data Science Pipeline	
Pipeline	Efficiency of the end-to-end data management process
Forward-looking	Plans for expanding data collection and improving utilisation

The goal of this benchmarking study was to provide a robust perspective on the current landscape of data science utilisation within innovation funding agencies. Anonymity was guaranteed, ensuring candid feedback and genuine insights. For a comprehensive breakdown of questions, or to participate in the benchmarking, interested innovation funding agencies may contact IGL at innovationgrowthlab@nesta.org.uk.

Agency benchmarking results

The benchmarking results across the categories in this framework immediately reveal the extent to which agencies have differing strengths and areas for development. While some excel in data collection and serving organisational use cases, others have focused on building skilled teams and investing in data pipelines. This diversity in the distribution highlights the potential for investing in data-driven decision making across agencies, and the opportunity for peer-to-peer learning between them to foster mutual growth and improvement across the board.
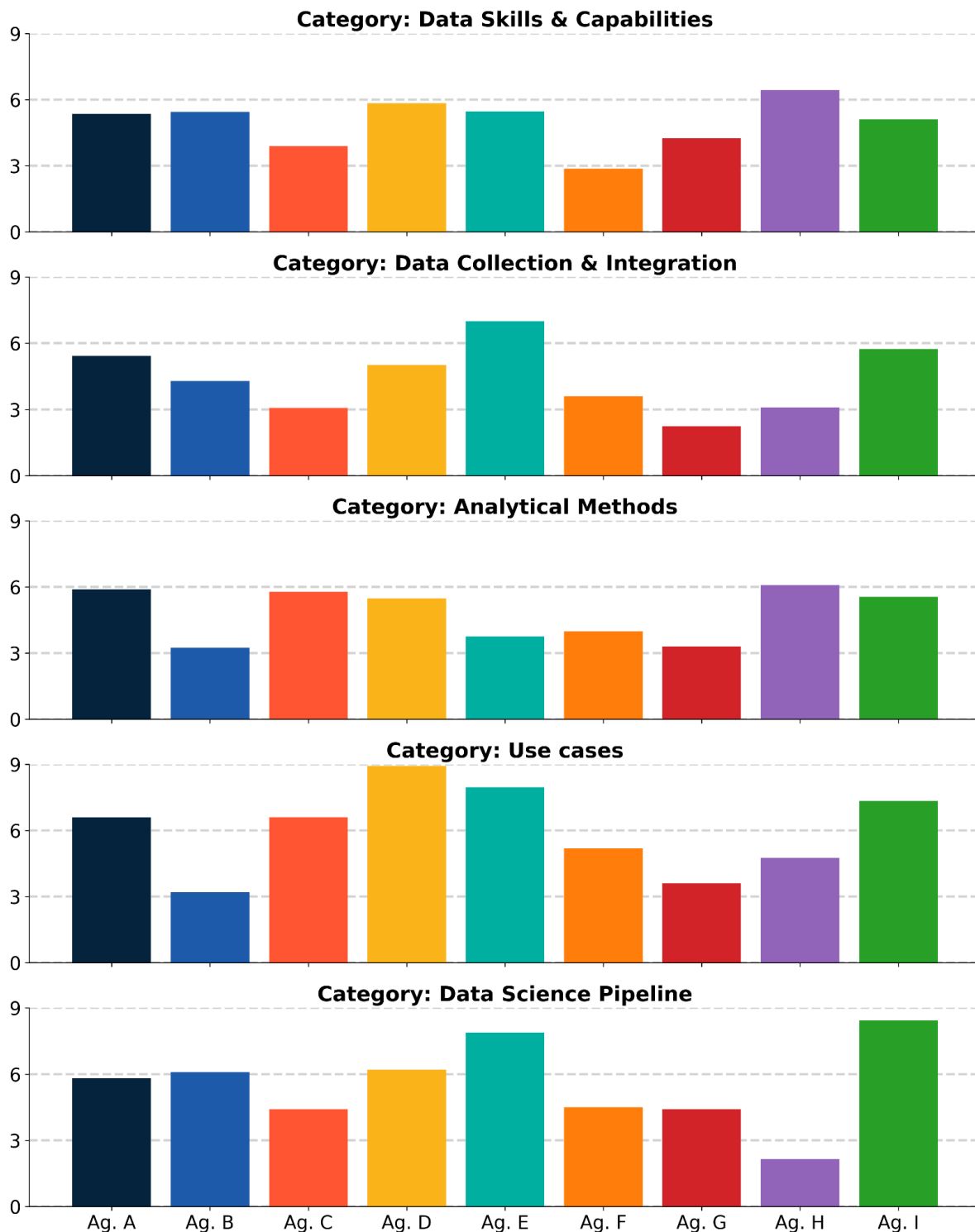
Not all areas show strong differences between agencies. Some categories, like **Data Skills & Capabilities** and **Data Science Pipelines**, show more consistent strength across agencies, contrasting with dimensions such as **Analytical Methods** and **Use Cases**, which exhibit significant variation.

Additionally, many dimensions remain far from the maximum score, even for the strongest performers, indicating that all agencies have dimensions to develop. Since these areas of expertise do not necessarily overlap, peer learning becomes especially valuable, as agencies can address gaps by learning from others' strengths.
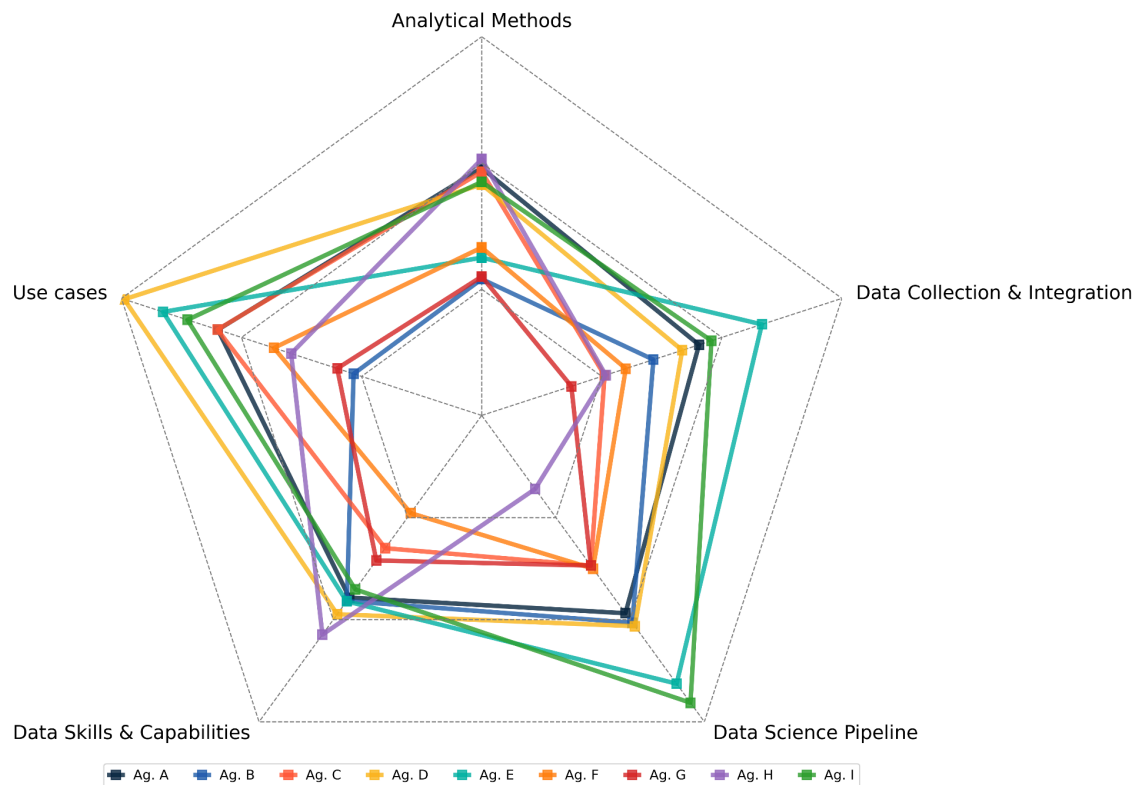
Data Skills & Capabilities emerged as one of the stronger areas across most agencies, though not without gaps. Agencies with more structured data teams and targeted skill-building efforts reported higher scores, while others struggled due to limitations in recruitment or the development of in-house expertise. The results suggest a clear path for agencies to improve by focusing on internal training and upskilling, as well as addressing recruitment bottlenecks, to ensure that data teams can support both current and future needs, and reduce reliance on external insight and data providers.

Data Collection & Integration scores varied significantly, with some agencies excelling due to their investment in centralised data systems, while others fell short because of fragmented, inconsistent data sources. Agencies that performed well in this area demonstrated clear strategies for managing data across different systems, ensuring consistency and accessibility. Those lagging can address these issues by prioritising the development of unified data collection systems and enhancing the technical infrastructure to improve the quality and reliability of their data.

The category **Analytical Methods** highlighted significant disparities in the methods deployed by agencies. A small number of agencies have adopted advanced techniques such as natural language processing and predictive analytics, allowing them to perform more sophisticated analyses. However, many others remain reliant on basic analytical methods, indicating a need for further investment in both tools and training. To close this gap, agencies should explore opportunities to build internal capacity for more advanced analytics, or nurturing long-term collaborations with policy-oriented academics.

Use Cases were another area where agencies demonstrated significant variation. While some agencies have effectively applied their data to a range of strategic and operational decisions, others have focused more narrowly on fulfilling immediate reporting requirements. The lower scores in this category suggest that many agencies are not yet fully capitalising on the potential of their data for broader decision-making. To achieve this, agencies require both a mindset shift about the purpose of collecting data, as well as an investment into new projects that test advanced applications for monitoring impact, maximising efficiency, and extracting insights.

Finally, **Data Science Pipelines** showed relatively strong performance, particularly among agencies that have made progress in refining their workflows and institutional processes for data use, from collection to analysis and dissemination. However, there is still room for improvement, especially for agencies that are in the earlier stages of building an integrated system for requesting and communicating insights from data. These agencies should focus on refining their processes to ensure insights flow smoothly across systems, including the institutional process that leads to evidence requests and the structure around the exploration of novel methods and data that could better meet the demand for insights.

The results across these categories highlight both strengths and areas for growth, but to fully understand the disparities and opportunities, it is important to examine the specific subcomponents within each category. By doing so, one can identify the underlying factors that contribute to the scores, such as the structure and expertise of data teams, the processes used for data collection and integration, the depth of analytical methods employed, and how effectively agencies are applying data in practice.

The following sections break down each category in detail, providing a closer look at the elements that influenced the benchmarking results.

Breakdown of benchmarking categories

1. Data Skills & Capabilities

This category evaluates the structure and expertise of data teams within agencies. It considers three main factors: Team Composition, Skills, and Capabilities. These subcomponents measure the size and specialisation of the teams, the range of technical and analytical skills available, and the ability to apply these skills effectively in real-world scenarios. This category provides insight into how well agencies are equipped to manage and utilise data for decision-making and strategic purposes.
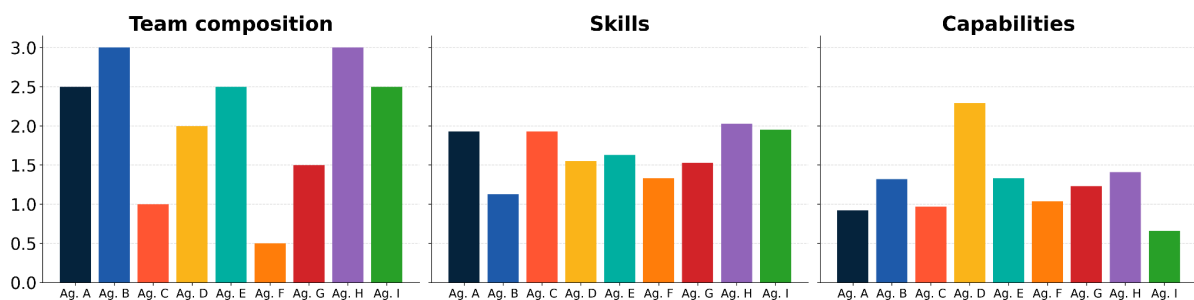

- **Team Composition:** Agencies with larger and more varied data teams performed better. They often have roles with specific responsibilities, allowing them to focus on different aspects of data work. Agencies with smaller teams or those with less specialisation showed lower results, indicating a need to build more well-rounded teams.

- **Skills:** Agencies that frequently applied descriptive statistics and inference methods scored higher. In contrast, more advanced techniques, like machine learning, predictive modelling, and network analysis, were rarely used or commissioned externally across the board. This suggests a trend where most agencies have strong foundational skills but rely on external expertise for more complex tasks.

- **Capabilities:** Scores are low across agencies due to the limited use of advanced data storage and computing resources. Most agencies rely on basic local storage and personal computers, with little use of cloud-based or high-performance setups. Expanding the use of advanced data techniques would drive the need for better infrastructure.

In summary, while agencies show strength in fundamental data skills and basic computing setups, gaps remain in advanced capabilities and data access. Addressing these gaps by expanding internal skills and upgrading infrastructure could significantly enhance their data-handling capacity.

2. Data Collection & Integration

Central to any data-driven organisation is its ability to gather and consolidate information. This category evaluates the breadth (covering various domains, sectors, or topics) and depth (granularity within each domain) of data collected. It also assesses the proficiency of the agency in merging and harmonising datasets from diverse sources and emphasises the quality of data in terms of accuracy, timeliness, and reliability.
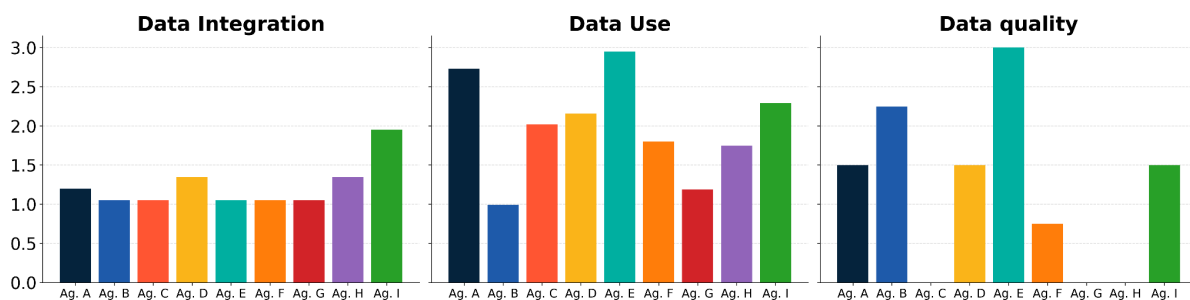

- **Data Integration:** Agencies that use more advanced methods like automated data import and integration tools would score higher, but most still rely on simpler methods like spreadsheet operations or manual data entry. This limits the integration of various data sources, creating fragmentation across departments.

- **Data Use:**

  - **Extensive Margin:** Agencies that use data for various purposes, such as selecting applicants, communication efforts, and feedback processes, scored higher. Some agencies plan to expand their data use but currently rely on fewer, more basic applications, with many reporting that more than half of their available data remains unused.

  - **Intensive Margin:** The complexity of data use also varies significantly, and correlates strongly with its extensive counterpart. Some agencies perform in-depth analysis using multiple sources and sophisticated methods, while others rely on more basic analysis, applying straightforward techniques to readily available data.

- **Data Quality:** Agencies with strong processes for ensuring data quality, such as harmonisation and rigorous checks, performed better. However, incomplete or missing data remains a common issue for many, and some only had a few, ad-hoc checks in place. Establishing better quality control measures and adopting advanced data cleaning techniques could enhance data reliability.

In summary, while agencies are making progress with data integration and use, more advanced methods and quality controls could significantly improve their ability to collect and utilise data effectively. The quality of data, in particular, was identified as a critical issue and was discussed at length during the Innovation Data Dialogues.

3. Analytical Methods

An agency's analytical approaches determine the quality of insights it can extract from data. This category examines the types of analyses performed by agencies and how advanced their data analysis techniques are. It looks at three main levels, ranging from foundational analytical techniques to sophisticated, cutting-edge data analysis methods. The objective is to assess the agency's analytical maturity and its readiness to meet the growing demand for evidence-based decision-making.
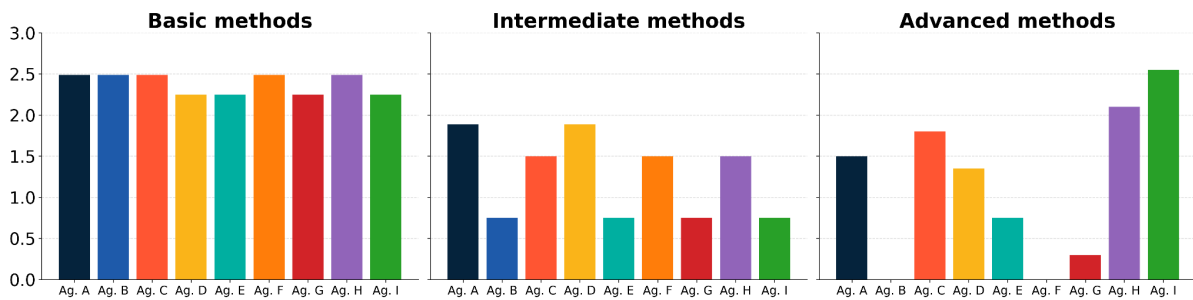

- **Basic Methods**: Agencies that frequently use descriptive statistics through visualisations, such as charts and dashboards, scored higher. These methods are employed multiple times across all agencies, showing that foundational analytical skills are well-established and widely utilised. However, inferential statistics, such as hypothesis testing, are less commonly used, with some agencies relying on external consultants for these.

- **Intermediate Methods:** Agencies that apply techniques like regression analysis internally scored higher, showing apt analytical capacity. However, many still outsource this work or have the skills but rarely apply them. Time series analysis is even more underutilised, with no agencies using it internally and only some commissioning it externally, which awards half points. This suggests that many are still developing their capacity to perform intermediate methods regularly.

- **Advanced Methods**: Agencies that use advanced techniques like machine learning, natural language processing, and predictive modelling scored the highest. However, only a few agencies have started to use these methods, with others either not using them or commissioning the work externally, receiving half points. Geospatial analysis and network analysis are also rare, with only few agencies applying these advanced techniques internally.

In summary, agencies generally demonstrate strong capabilities in basic methods, but the use of intermediate and advanced techniques is less common. Expanding internal expertise in these areas could help agencies reduce reliance on external consultants and improve their ability to handle complex analyses.

4. Use Cases

Beyond mere data analysis, the real value lies in actioning those insights. This category assesses how agencies use data for various purposes, focusing on Impact, Efficiency, and Insight. Each subcomponent evaluates how well agencies leverage data to improve decision-making, streamline processes, and gain deeper insights into their operations and impact.
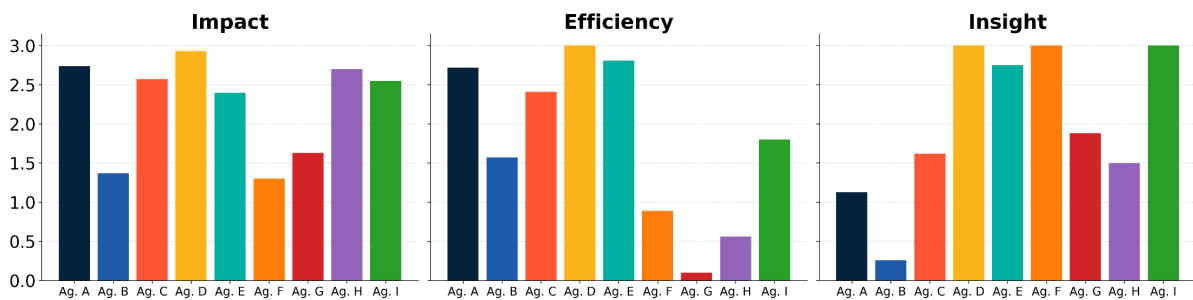

- **Impact**: Agencies which frequently use data for risk mitigation, strategic planning, and impact evaluation scored higher. Regular use of data to evaluate funding outcomes and anticipate risks enables agencies to make better-informed decisions. However, some agencies only engage in these analyses occasionally or not at all, indicating room for improvement in using data to drive strategic planning and assess long-term impact.

- **Efficiency**: Agencies that use data to track applicant demographics, optimise processes, and assess equity, diversity, and inclusion performed well in this area. By using data to streamline operations and improve fairness in their processes, these agencies improve both the applicant experience and internal workflows. Others are still in the early stages of applying data to these ends, highlighting an opportunity to enhance operational efficiency through better data use.

- **Insight**: Data is used to gain deeper insights into the innovation ecosystem, map clusters, and ensure policy alignment. Agencies that regularly perform predictive analysis and innovation ecosystem mapping scored higher, reflecting their ability to anticipate future trends and align their efforts with broader policy objectives. Some agencies plan to expand their data use in these areas, but there is still significant variation in how frequently these analyses are conducted.

In summary, while many agencies are making strides in applying data for impact, efficiency, and insight, others are still developing their capabilities. Expanding the use of data-driven insights, particularly in strategic planning and predictive analysis, would enhance decision-making and improve overall outcomes.

5. Data Science Pipeline

This category captures the entire lifecycle of data management within agencies, focusing on the robustness and efficiency of existing processes, as well as forward-looking strategies for improvement. It is divided into Pipeline, which assesses current practices, and Forward, which evaluates agencies' future plans for expanding data capabilities.
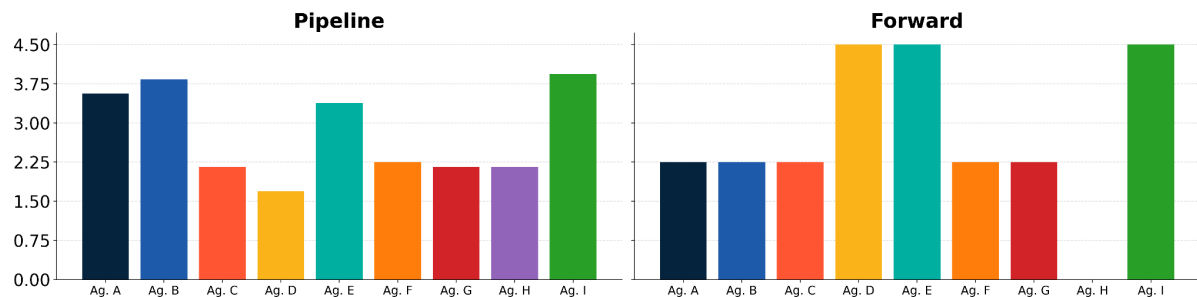
- **Pipeline**: Most agencies have structured processes for receiving data requests and setting clear timelines, ensuring that data is used effectively. Agencies with more advanced communication of results and clear top-down and bottom-up request channels scored higher. However, some agencies reported that a significant portion of their data remains unused in decision-making, weighing down on their scores and indicating the need for better integration of data into everyday operations.

- **Forward**: When it comes to forward-looking strategies, agencies vary quite a bit. While some actively collect new data to inform strategic decisions, others are less proactive. Common wishes include access to more longitudinal data, better integration of data on companies, and improved access to external data sources. Agencies that are already planning or engaging in extensive data collection efforts are better positioned to enhance their data-driven decision-making in the future.

In summary, agencies are making progress in structuring their data pipelines, but there is still potential to improve data utilisation and expand strategic data collection efforts for future decision-making.

Strategic Implications

The findings of this benchmarking exercise highlight several key areas where agencies can improve their use of data for decision-making and strategic operations. Based on the insights gathered, the following strategic implications and recommendations can guide agencies in their future efforts:

1. **Enhancing data literacy and internal capacity building**
   While many agencies have the foundational skills and resources, their full potential is often untapped. Agencies should focus on improving internal data literacy for both technical teams and decision-makers, ensuring they can understand and act on data-driven insights. Upskilling programmes can also reduce reliance on external consultants and enhance in-house capabilities.

2. **Improving data integration and breaking down silos**
   Internal data silos continue to be a significant challenge. Agencies should develop unified data management strategies that promote sharing and integration across departments. This could involve centralised data warehousing supported by embedded data analysts to balance accessibility with departmental needs. Clear communication channels are also essential.

3. **Focusing on data quality and governance**
   Ensuring data accuracy, timeliness, and consistency is critical for actionable insights. Agencies should establish more rigorous data quality control measures, with dedicated staff responsible for data cleaning and validation. Clear data governance frameworks are also essential to maintain high standards, reduce errors, and ensure data reliability.

4. **Expanding analytical capabilities**
   Foundational data analysis methods are widely used, but advanced techniques remain underutilised. Agencies should explore ways to build in-house expertise, whether through staff training, partnerships with academic institutions, or collaborations with external experts. Expanding analytical tools will help extract more valuable insights.

5. **Supporting continued experimentation and peer learning**
   Organising data systems is only the first step. The real value lies in using this data to draw insights and explore new strategies. Experimentation offers agencies a practical way to test ideas, assess their impact, and improve operations. In parallel, peer learning, such as through the [Innovation Data Dialogues](#), enables agencies to share experiences and collaborate on solutions to common problems.

By addressing these priorities, agencies can unlock the full potential of their data, foster more informed decision-making, and enhance their overall impact on innovation funding.