Data-Driven Dilemmas in Innovation Funding

14 questions about unlocking access, managing capabilities, and structuring for success

This brief summarises key discussions and challenges from the Innovation Data Dialogues, a series of collaborative sessions organised by the Innovation Growth Lab with innovation funding agencies, focused on data governance, exploitation, and experimentation as part of a project with the European Innovation Council. The discussions highlighted three recurring challenges: ensuring timely access to the right data, balancing internal and external capabilities, and building an organisational structure that supports sustainable success. These challenges are explored here through fourteen guiding questions that provide a deeper understanding of the issues discussed.

September 2024

Authors:
Sara García Arteagoitia, David Ampudia Vicente, George Richardson

The Innovation Data Dialogues

The Innovation Data Dialogues (IDDs), supported by the European Innovation Council (EIC) as part of the project "Driving Strategic Data-Driven Experimentation", were a series of collaborative sessions between twelve regional, national and supranational innovation funding agencies organised by the Innovation Growth Lab during the spring of 2024 to explore critical challenges and opportunities in data-driven innovation. They were part of an overall project aimed at providing evidence-based advice on data-driven, experimental, and collaborative public sector innovation processes, to support the development of strategic intelligence at the European Innovation Council.

Before joining the sessions, the agencies took part in a benchmarking exercise to assess how they use data in their daily work. The ensuing reports compared the agencies' data skills and capabilities, data collection and integration processes, the analytical methods used for data exploitation, the range of data use cases, and their data science pipelines. This exercise helped agencies understand their standing and identify possible blind spots, setting the stage for sharing best practices and discussing common challenges at the Innovation Data Dialogues.

The IDDs addressed three dimensions of data-driven innovation funding. The first dialogue focused on data governance, discussing its fundamental concepts, ownership structures, and sharing best practices among agencies. The second session explored frameworks for data-enabled policy, with a particular focus on how these frameworks can be put into practice through a series of case studies and discussions. The final session highlighted experimental methods for innovation funding, comparing traditional evaluations with experimental approaches and showcasing current examples from partner organisations.

Setting up the Innovation Data Dialogues: Why data-driven innovation funding?

A data-driven organisation routinely collects, stores, and analyses high-quality data to support a range of operational and strategic decisions. An effective data function will combine engineering and analytical skills to develop valuable data pipelines and insights. Pipelines that collect and process data across funding programmes allow routines to be automated, reducing workload and making processes easier for applicants, assessors and employees alike. They also ensure the accuracy and insightfulness of data used for downstream decision-making.

Moreover, data pipelines create a robust institutional memory, which is crucial for agencies that need to evaluate past performance, benchmark their programmes to external standards, and coordinate with other funding organisations. This ability to look back and measure impact supports continuous monitoring and helps agencies learn from their successes and mistakes. It also allows for the standardisation of performance metrics, making it easier to spot trends, address gaps, and ensure that programmes remain aligned with broader policy goals.

Beyond internal benefits, effective data use supports broader research and innovation. By strengthening existing analysis pipelines, agencies can contribute valuable insights to the wider community, driving forward knowledge and advancing public sector practices. Developing robust data-sharing frameworks can also pay large dividends. When agencies share data, they enable researchers, policymakers, and other stakeholders to explore new ideas, test hypotheses, and develop evidence-based solutions that address complex challenges, not only improving their own practices, but also playing a vital role in shaping a more informed and effective public sector.

Data-driven approaches also open the door to experimentation. With a solid data foundation, agencies can test new methods, pilot innovative approaches, and refine their models based on feedback and results. This experimental mindset is essential in a rapidly changing environment where agencies must adapt effectively, allowing for quick testing but also accelerating learning by rapidly generating evidence that can be disseminated across the organisation. This enables agencies to quickly identify what works, cut losses on approaches that do not deliver, and scale up successful strategies when positive outcomes are observed.

Three challenges and fourteen questions

Over the course of three months, twelve innovation funding agencies participated in the Innovation Data Dialogues, sharing common difficulties and exchanging best practices. These fruitful discussions yielded three main challenges and fourteen questions around unlocking access to data, managing capabilities and structuring organisations, summarised below. The questions reflect the discussions that occurred across the three Innovation Data Dialogues and cover the full spectrum of owning and understanding the value of data assets through the capabilities needed to extract most value from them up to organising those capabilities to do this most effectively.

Challenge 1: Accessing the right data at the right moment in the right format

Data access was identified throughout the IDDs as one of the most consistent and limiting challenges in innovation funding. This challenge encompasses different dimensions ranging from identifying needed data through accessing data to integrating data sources. As the most fundamental challenge, it poses the greatest limitation to data-driven innovation.

1. What data is available?

Agencies often lack real-time visibility into available data across departments, which hampers efficient data use and sharing. The knowledge about what data is available is seldom structured and accessible to all actors in the organisation, and often sits dispersed across departments and only in the minds of those who regularly use the data. Lacking a centralised data catalogue hinders starting a search for data and limits the ability to share data with third parties for research, analytical or comparison purposes, as the legal standards require that data be identified before a sharing request can be approved. Data catalogues are a necessary first step in laying the foundation for smooth access to data. However, the effort to create data catalogues comes with its own challenges. Despite the existence of commercial solutions, no shared standards exist yet across innovation funding agencies that could be easily used, reproduced and shared.

2. What data can be used?

It is not always clear that past data, as useful as it might be to answer operational and strategic questions, can be used to improve decision-making in compliance with General Data Protection Regulation (GDPR). GDPR restricts data storage and use to purposes closely related to the original reason for data collection unless further processing for compatible purposes or exceptions for scientific research can be justified. Therefore, using historical data for research or cross-project analysis may not always be allowed without appropriate safeguards or obtaining explicit consent. This problem can sometimes be mitigated by adding new clauses to the data collection process during the application stage, ensuring that individuals are aware of and agree to broader uses of their data for research purposes.

3. Who owns the data?

The problem of data ownership is also extended across organisations. Data is accumulated over time as a consequence of distinct programmes without coordinated

data-collection efforts. It is then unclear who has the mandate to maintain, use, update, integrate or share the data, who can make decisions about the future of the data and who can answer questions about the data. Siloed operations, unclear data governance and different personnel structures make it difficult to know who can theoretically and practically access the data once it has been located.

4. How good is the data?

Once the data is accessed, there is a question of data quality - it needs to be accurate, complete, consistent, valid, updated, unique and traceable. Robust data infrastructure is required to achieve this. However, without a systematic review of data and a shared organisational agreement on how errors and inconsistencies will be solved, the solutions end up being ad hoc and might differ considerably across projects, leading to non-comparable analyses. It also makes it harder to match and merge datasets, limiting downstream analysis opportunities.

5. How is data stored?

Inconsistent data storage is a common problem derived from uncoordinated data collection. Each department might use different software and formats to store the data and have different rules about how the data may be retrieved (e.g. have a cap on how much data can be downloaded at a time). Often the requested solution for issues of data identification, quality, ownership and access is the use of a centralised data warehouse, where data is stored, audited and curated for the organisation's use. This requires senior buy-in, cross-organisational effort, and resources to set up and maintain, which need to be weighed against the inefficiencies, frustrations, and limitations that more decentralised systems come with.

6. How is data accessed?

Centralised data is not enough to guarantee efficient access to data for all who need it. A centralised warehouse also needs to come with clear specifications of data access according to different departmental needs and abilities, ensuring that its existence facilitates more efficient access to high-quality data. Tiered access to more raw or more curated data based on departmental needs and analytical abilities is a way in which some organisations solve this issue.

7. How is data presented?

Dashboards are often a convenient response to this question, albeit a frequently problematic one. At first sight, dashboards seem like a good solution to present data as

they transform raw data into visuals with the potential for high-level summaries, interactive data exploration, trend analysis, and reporting. They are also a tangible output for contracted work, making them attractive for commissioning. However, to be useful tools, dashboards must be set up in specific ways that deal with innovation funding data: where data has been collected across a patchwork of funding programmes, the dashboard interface must make limitations and inconsistencies in the data clear to the user; users must understand the meaning and provenance of variables presented to them to draw the correct conclusions; and crucially, dashboards require ongoing maintenance (which is harder to contract out), lest they become obsolete quickly. Dashboards also cannot be used to produce complex analyses and deep or highly detailed insights and are ill-suited for one-time analysis. Therefore, they should only be used as a decision-supporting tool when they can be set up to match analytical needs; otherwise, other methods may be more suitable for presenting data and supporting decision-making, ranging from traditional and less resource-intensive reporting methods to automated statistical analyses, or tools that facilitate human-AI collaboration to yield bespoke insights.

Challenge 2: Finding the right mix of internal and commissioned capabilities

Innovation agencies face an ever-changing landscape of policy priorities and technologies that they have to address with the development of new programmes or the continuation or tweaking of existing ones. Hence, the set of questions they need to answer requires a mix of known and new tasks. Being able to ensure and maintain access to the right data capabilities to face these changing needs was identified throughout the IDDs as a recurring challenge for numerous agencies. This challenge often requires deciding when and how to grow a data team, contract support to expand or enhance internal capacities or commission projects out completely.

1. What is the right composition of skills?

Data team sizes vary as a function of organisational size, but that does not necessarily mean that the diversity of capabilities also grows proportionally. Often profiles are repeated to respond to more demand for the same type of (reporting) activities and are based on traditional structures, with some profiles being more policy-oriented and others more data-oriented. The teams are thus often designed to address repeating needs considered part of the core function. Over time, new data needs - analytical or infrastructural - can emerge that the teams might struggle to address due to limited capacity or capability. These gaps might then be filled through hiring, training and team reconfiguration or bringing in external support.

2. When does commissioning bring value?

The appetite for newer, more specialised analysis often leads to the chicken-and-egg question of whether such specialised analysis should precede or follow the hiring of more specialised skills. This issue is often solved by contracting the task out since this approach gives access to state-of-the-art knowledge, reduces upskilling needs and ensures clear deliverables. Hence, more in-depth, ad hoc analyses often end up being contracted out as part of pilot projects.

3. What projects can be commissioned?

Contracting out requires that a certain set of pre-conditions be met. It first requires the right contracting capabilities: to contract out specialised skills, the contracting organisation needs to be able to understand and evaluate the quality of bids and the results and serve as an apt counterpart in the research or development process. The project also needs to lend itself to clear deliverables or be adaptable, to match more dynamic goals, as is the case with exploratory projects. The projects ideally also need to be finite, since maintenance-intensive projects - such as dashboards - will either require ongoing contracting or the development of internal roles that can engage in maintenance and enhancements.

4. How to choose the right type of solution?

Contracting out also poses the question of choosing between off-the-shelf and boutique solutions. The former allows for rapid delivery in theory but might be harder to align with evolving policy and programme needs, is more susceptible to product discontinuation, and could end up creating parallel systems. More boutique solutions will be able to adapt to evolving needs but require an ongoing relationship with an external partner as well as a maintenance and updating contract over numerous budget cycles. In both cases, the use of open protocols, open source data formats, and established standards can provide useful flexibility.

5. When should capacity and capabilities be built internally?

The more central the contracted task is to strategic decision-making or operational capacity, the more the development of internal capacity should be considered. If an analysis is repeatedly outsourced, it likely signals a need to build this capacity internally. However, developing internal capacity comes at different costs for different organisations. For some, where hiring external top talent into public administration is relatively quick and easy, adapting the team to evolving challenges through hiring is an accessible solution. But they may still face challenges hiring people with relevant domain

expertise and data people hired early on into a data-immature organisation might face difficulties nonetheless. For other organisations, where hiring is complex and lengthy, the pool of candidates is limited, or growing the organisation's core budget is much harder than finding room in the flexible budget, setting up an external contract might be the best (or only) solution. The latter becomes a risky strategy when what is contracted out is important for the core function of the organisation.

Challenge 3: Designing the right organisational structure for sustainable success

Choosing the right organisational structure in an ever-changing environment is one of the crucial challenges organisations face in solving their issues. These decisions include how the organisation organises both vertically and horizontally and how it makes use of its resources and abilities most efficiently and effectively.

1. Where should the data team(s) sit?

A centralised data team can enhance operational consistency, ensure data compliance, and align its efforts with the organisation's overarching strategy; but may struggle to keep up to date with the different departmental contexts, reduce data literacy across the organisation, and turn into a service centre instead of promoting wider data foundations in the organisation. Decentralised teams can be better embedded in departmental priorities and have quicker response times, but will have a harder time maintaining equal data quality and governance standards across the organisation and might duplicate efforts in some tasks. Hybrid models, where a central team sets and maintains data standards, regulates access to data and enforces data governance while data analysts are embedded in the departments, work as a compromise to achieve high data standards and easy access while delivering on the departmental needs and ensuring widespread data literacy. This model requires clear communication channels and data-sharing mandates to avoid miscommunication and barriers to data access.

2. Who sets the data strategy?

Often there is a mismatch between the demand and supply side of data projects. Data analysts might be able to offer more diverse analysis than is being asked of them but lack a clear sense of how to deploy novel capabilities to meet policy priorities. Policymakers have a clear sense of strategic priorities but might not understand the possibilities available to them through data-driven analysis. The most productive outcomes are achieved when dialogue establishes trust, and that trust leads to solutions that most fit immediate business needs and room for exploration of new tools and

methods. In practice, data analysts have often started bottom-up initiatives to address existing challenges in new ways, but limited room and resources for exploration and experimentation halt the effect of these initiatives or result in solutions that are not well aligned with organisational needs.

Outlook

The Innovation Data Dialogues made clear that many data challenges are shared among innovation funding organisations, but that there are few one-size-fits-all approaches to solving the issues. The discussions identified challenges of creating and accessing high-value data, building and using the most useful capabilities, and organisational design for success. While the structures and objectives of each organisation direct and limit how challenges can be addressed in their respective contexts, four key response areas have also emerged.

One of these is developing an understanding of how exactly becoming a data-driven organisation will create value and establishing a matching data strategy. Developing this kind of theory of change will objectively prioritise business needs that an institution must address and assess which ones will benefit from more quantitative information being made available to support strategic decision making. At the organisational level, such an exercise will also help to answer structural issues concerning where data teams should sit and how their priorities should be set.

While strategies provide direction and structure for an organisation, delivering on intentions usually requires solid infrastructure. For work involving data, the foundations of data governance, access, and management are central pillars. Fortunately, there are well-established technological solutions and social processes which can be deployed to address each of these. The greater challenge stems from the fragmentation of responsibility for data and associated value chains across teams and departments. A process that brings together internal stakeholders around overlapping interests, with sponsorship from senior leadership and neutral facilitation, is one answer to road-mapping and delivering these types of foundations.

Attaining the right mix of internal capabilities and using external capabilities appropriately is another area that can unlock significant value. Although the role of data scientists became popularised over the last decade, it is not always necessarily the most critical role an organisation should consider, especially when resources are limited. Data engineers are critical when automation of routine analyses and data enrichment is required, whereas analysts often have statistics knowledge that is needed to provide more causal analysis. Outsourcing is often useful for R&D and ‘test-and-learn’ initiatives. In all cases, there can be a significant return on investment by empowering a technical delivery manager to oversee scope, interoperability and alignment.

It is also clear from the IDDs that innovation funding agencies have much to learn from each other. The continuation of the dialogues would allow peer exchange to continue. Another more concrete collaborative effort might be the formation of task forces, where agencies pool resources to develop methods and build tools to address common analytical challenges. These might cover approaches to forecasting, evaluating programme impact and linking data to open data sources, or finding ways to share data responsibly.

These areas are ones which currently present strong opportunities to build the foundations of the sector, based on the IDDs. It is important to note that each of these can be adjusted to fit a specific organisation's needs, size and capacity and that the order they have been written here does not reflect the specific order in which they should be addressed.

Innovation funding organisations are not unique in seeking to become more data-driven, but the domain that they operate within and their rhythms and processes present a need for specialised approaches. The challenges within innovation funding organisations are also ever-evolving, responding to policy needs and a dynamic ecosystem. The path forward will be built on strong foundations for using data, exploration and experimentation with new approaches, and institutionally embedded openness to share the lessons learned.

The **Innovation Growth Lab (IGL)** is a global policy lab that helps governments develop more effective policies to increase innovation and productivity. IGL's mission is to foster productive, sustainable, and inclusive economies through novel policy ideas, experimentation, data and evidence. IGL works with policymakers, researchers, practitioners and funders to address key policy challenges in the fields of science, innovation, entrepreneurship and business policies.