FINAL

Data Pitch evaluation

Report to

data-pitch

INNOVATION PROGRAMME

LE London Economics

December 2019

About London Economics

London Economics is one of Europe's leading specialist economics and policy consultancies. Based in London and with offices and associate offices in five other European capitals, we advise an international client base throughout Europe and beyond on economic and financial analysis, litigation support, policy development and evaluation, business strategy, and regulatory and competition policy.

Our consultants are highly-qualified economists who apply a wide range of analytical tools to tackle complex problems across the business and policy spheres. Our approach combines the use of economic theory and sophisticated quantitative methods, including the latest insights from behavioural economics, with practical know-how ranging from commonly used market research tools to advanced experimental methods at the frontier of applied social science.

We are committed to providing customer service to world-class standards and take pride in our clients' success. For more information, please visit www.londoneconomics.co.uk.

Head Office: Somerset House, New Wing, Strand, London, WC2R 1LA, United Kingdom.

w: londoneconomics.co.uk e: info@londoneconomics.co.uk : @LondonEconomics
t: +44 (0)20 3701 7700 f: +44 (0)20 3701 7701

Authors

Moritz Godel, Wouter Landzaat, Ryan Perkins

Table of Contents

Table of Contents

Executive summary

The Data Pitch programme

Data Pitch is a pioneering programme aimed at facilitating open innovation through data-sharing. Data Pitch helped data providers to develop open innovation challenges and facilitated the matching between data providers and start-ups. Data Pitch establishes a transnational data innovation ecosystem that creates collaboration between data providers on the one hand, and start-ups with fresh ideas for data-driven products and services on the other.

Data Pitch combines a thorough preparation phase to identify suitable challenges and an acceleration phase during which participating start-ups are supported through:

- A grant (equity-free investment) of up to €100,000;
- introductions to investors;
- 6-month business accelerator programme with ODI and Beta-I;
- peer-networking and support via meetups;
- access to training materials and webinars by Data Pitch experts;
- legal counsel for IP protection;
- drafting of a data sharing agreement between Data Pitch partners.

The key distinguishing feature of Data Pitch is that it matched data providers that possessed data that they wanted to use for open innovation, with start-ups with the wherewithal to exploit this data. Start-ups in the "provider" challenges were matched by the Data Pitch consortium, whereas start-ups in the "sector" or "open" challenges had to find their own data provider. All participants had to have a partnered data provider.

This study

The study explores and evaluates the impact of Data Pitch on participants; participants include both data providers and start-ups. The study uses a mix of quantitative and qualitative data obtained from programme documentation, published sources, and primary data collection.

The main quantitative data collection for this impact assessment was a survey of the 47 start-ups funded through Data Pitch (successful applicants). 41 observations were collected. This survey was complemented with a survey of applicants who were not successful (9 observations).

Qualitative information was collected through interviews with start-ups and data providers between September and November, 2019.

Additional information was obtained from interviews with the Data Pitch consortium, programme documentation, and external data sources. Six case studies illustrate the achievements of selected start-ups in the provider and sectoral challenges.

Participants

Over the two Data Pitch calls, 47 start-ups, out of a total number of 239 applicants, were successful in gaining admission to the programme. 22 of the different challenges were matched with start-ups, with one Data Provider challenge in each call left unmatched. 18 start-ups were selected for funding

in Call 1 and this increased to 27 in Call 2. Thirteen data providers participated in the programme (5 in Call 1 and 8 in Call 2).

Many of the participating start-ups worked across sectors, with the aim to enter new sectors being a key motivator for participating in Data Pitch. A concentration of participating start-ups can be found in the health, financial and transport/mobility sectors.

Programme impacts

During the Data Pitch programme, firms increased their sales by a mean of €36,554, received investments of a mean of €82,448, realised efficiencies of €17,168 and increased their employment by an average of 2 employees. On average, start-ups generated €599,432 in sales and €338,862 in investment per GB of data shared with them through Data Pitch.

Both ROI and leveraged investment are already substantial during the programme. By the end of the programme, the total Data Pitch resources already attracted 50% equivalent value from other investment opportunities. One year after the end of the programme, ROI already exceeded 100% and leveraged investment was already approaching 300%.

Table 1 Return on Investment and leveraged investment; realised and projected figures

	During the programme	6 months following the programme	12 months following the programme	Projected annual figure by 2022
Return on Investment	23%	91%	103%	459%
Leveraged investment	50%	82%	278%	N/A

Note: data on 6 and 12 months after the programme are based on data from Cohort 1 only. The figures account for this by adjusting the investment provided through Data Pitch based on the number of start-ups for which data is available.

Source: Bi-weekly monitoring updates, 6 months progress update, 12 months progress update, revenue growth projection

The majority of start-ups would not have been able to access the same, or similar, data without Data Pitch; this is especially true for start-ups working in the financial and medical sectors. Access to data does seem to influence the ability of start-ups to attract additional funding. Start-ups that could access data outside of Data Pitch attracted, on average, €141,000 more in additional funding than start-ups that would not be able to access data. Start-ups typically had full control over the data when building the solution. Similarly, data was typically stored under the control of the start-ups.

The majority of start-ups used machine learning in their solution. The outcomes tracked during the programme provide some evidence that using machine learning helps attract investment. There is an impressive difference between start-ups that use machine learning and those that do not. Start-ups that did use machine learning attracted, on average, €108,000 more in investment than those that did not use machine learning.

Regarding growth opportunities, start-ups in the sector challenges had higher growth expectations for their product. This (perceived) ability to scale a solution has a distinct impact on the ability to attract funding; with easier perceived ability to scale being associated with increases in additional funding received.

Successful applicants received, on average, more external funding than unsuccessful applicants (not including the €100,000 received through Data Pitch). Some applicants, both successful and unsuccessful, were able to attract substantial investments upwards of €500,000.

Projecting performance of funded start-ups into the future, a forecast model predicts that average revenue for start-ups will grow from €147,723 in 2019, to €833,555 in 2022. Combined with the success/failure rate of businesses, this implies a growth of total annual revenue to €35,784,385 from €6,896,000. This is equivalent to a growth of 73% per year up until 2022.

There is considerable variation between outcomes for start-ups operating in the same sector. Tentatively, it seems that health start-ups still rely on investment, start-ups working in heavy industry are maturing to be sales-driven and start-ups in the financial sector show most employment growth.

Use of data

Data Pitch enabled start-ups to access data they would otherwise not have been able to access. Through the data provider challenges Data Pitch opened up data from data providers that would not be available for use by start-ups otherwise.

The main objectives of start-ups in Data Pitch was to make predictions with or identify patterns in data. The majority used machine learning to do this, especially start-ups in the sector challenges. Start-ups rated their solutions as moderately unique and innovative.

Impact on data providers

The impact on data providers has been reported mostly in terms of promoting the open innovation approach within the provider organisations. Data providers saw themselves participating in open innovation in the future and several felt they had absorbed sufficient knowledge to do so without external help. Several remarked especially that they were now further along in their journey to make their data more accessible and understandable for third parties. Some data providers mentioned that more contact with other data providers would have been useful. Creating a network of data providers that continues after the acceleration phase might be a useful extension of a programme like Data Pitch.

Although Data Providers went into Data Pitch with a business challenge, they saw Data Pitch as a learning opportunity regarding Open Innovation and data sharing. As such, measuring quantitative impacts were not prioritised and generally not available. However, one data provider estimated that the solution developed with their data in Data Pitch reduced the cost of a particular business process by 35%.

European collaboration

Data Pitch reached start-ups based in 13 countries, with cohort 2 broadening the reach of the programme. The United Kingdom represented a large proportion of all participants. The UK, together with Spain, also provided the highest number of participants invited for interviews.

24 out of 47 (51%) start-ups were partnered with a data provider located in a different country. In the data provider challenge, all but one of the 18 partnerships (94%) was a European collaboration.

Longer term impacts

Taking into account factors that may decrease the net positive impact (programme 'additionality') of Data Pitch, we conclude that the impact of Data Pitch is likely to be strictly positive, especially once longer-term impacts are taken into account. Data Pitch laid the groundwork for ongoing engagement in open innovation by some major European data providers, who appear ready to pursue their own initiatives going forward; Data Pitch has therefore acted as a demonstrator for data-driven open innovation.

The main benefit of Data Pitch for data providers was the learning experience from the programme. It is therefore likely that more tangible impacts as a result of participating in data-driven open innovation will only emerge over time.

Recommendations

- Sector challenges and provider challenges seem to have worked differently. For example, start-ups in the data providers challenges and the sector challenges differed in the methodologies used (sector challenge start-ups were more likely to use machine learning), the type of data used (sector challenge start-ups used video data, whereas data provider challenge start-ups did not) and the way data was stored (data provider challenge start-ups mostly files, whereas sector challenge start-ups mostly used relational databases) as shown in section 3.6. A more exclusive focus on matched challenges is likely to produce greater benefits and provides a better testbed for the hypothesis that reducing frictions inhibiting data sharing facilitates can unlock data-driven open innovation. The experience for participants in the sector challenges more closely resembled that of a standard accelerator, where the start-up is able to source data from external parties independently.
- More resources to prepare and connect data providers may maximise the programme impact. This could involve a "Phase 0" with a selection process and an 'acceleration' period focused on data providers to prepare them for working with start-ups on their challenges. The long-term benefits for data-driven open innovation could be cemented by the creation of a network of data providers that persists after the end of the programme.
- Despite the comprehensive support provided to the start-ups by Data Pitch, the more mature start-ups seem to have performed better. A clearer focus on start-ups with 'acceleration-stage' maturity (proven ability to deliver an MVP) may enhance the overall impact. Partly this could reflect the fact that more experienced start-ups are better able to select appropriate challenges both based on their technical viability, but also strategically, in terms of the applicability of an existing solution to a new market or industry. In this regard, a stricter separation between 'incubation-stage' start-ups, who may not be able to produce a working prototype by the end of the programme, and the more mature start-ups may be contemplated, so as to provide each type of start-up with the optimal support package, with more strategic advice being more appropriate for the more mature start-ups.
- There is some evidence that the impact of Data Pitch was stronger in sectors with higher barriers to data sharing (such as healthcare and finance, which data subjects see as particularly sensitive). For example, as discussed in section 3.7.2, start-ups in healthcare and finance were more likely to note that, without Data Pitch, they would not be able to access data. An ex-ante focus on such sectors may increase benefits. The selection of sectors should take into account their specific barriers to, and enablers of data sharing. For example, data-driven open innovation in the finance sector is supported by strong regulatory action (Open Banking, PSD2).

- The programme design and setup should facilitate robust evaluation of the programme itself. This includes having a clear evaluation strategy with well-defined success metrics for the programme, and comprehensive baseline data collection. Making long-term data sharing obligatory for participants may be considered. Which data is shared long-term, which may be a couple of years, depends on the success metric targeted by the programme, but may include revenues and employment figures of participants.
- More broadly, investment in further pilot projects is needed to develop the open innovation model and to find out what works. Parameters such as the selection of start-ups and type of support given should be comparatively analysed. An opportunity exists with using other European incubators, such as the European Data Incubator. Other European programmes may be used to experiment with, for instance, on-boarding processes in similar, but not identical, circumstances.

1 Background & context

1.1 Realising the value of data through openness

The ability to exploit data underpins more and more of the economy. Data drives innovation and productivity improvements for organisations that collect or control it, but also through its use by other parties. Often, these put data to innovative, unforeseen uses and create new applications that involve novel combinations of data from different sources.

The key to a productive data ecosystem is to enable reuse of data and integration with other datasets in a data value chain, which can include open, closed, and shared datasets. This means rethinking access to data and creating opportunities for data owners to work with other organisations, in particular innovative start-ups and SMEs that bring the creativity to use data in new ways to create value-added products and services.

The key objective of Data Pitch is to unlock the potential of data to solve critical challenges for industry, public institutions, individuals, and society as a whole by matching data holders with innovative start-ups and SMEs. Data Pitch thus addresses one of the most fundamental challenges faced by the data economy: How can data be used where it is most useful, and where it will generate the greatest positive impact.

Traditionally, most data tend to be locked in organisational 'silos'[1] – limiting the scope of what the data can be used for, and by whom (Ctrl-Shift, 2018). Data Pitch, by creating mutually beneficial data partnerships, and supporting the innovative businesses that can exploit them, can help to engender a multitude of economic and social benefits.

In economic terms, benefits come from two sources: a reduction in the cost of accessing data, and the ability to combine data from different sources. This in turn leads to:

- external benefits arising from increased use of data (i.e., more data being available, which may have value that is not reflected in the benefit received by the organisations and data subjects involved in the exchange);
- Higher productivity (i.e., making it easier to combine data from different sources lowers the cost of producing data-enabled products and services; access to bigger and richer datasets allows more precise predictions through better trained algorithms, etc.); and
- innovation in the form of new products and services from combining data in new ways across organisations and industry 'silos'.

Access to data can be very costly, especially for new-to-market SMEs[2]. Simply identifying relevant datasets, identifying their owners and negotiating access can be prohibitively difficult and resource-intensive. For participating start-ups, accessing data through Data Pitch partnership to avoid such costs and reduces the overall cost of data usage, allowing them to improve their technology and to produce more (and better) output.

[1] Data Silos refers to situations where data rarely leaves the organisation that collects it and is rarely used for purposes other than what it was first collected for.

[2] The Data Pitch programme was aimed at start-ups and SMEs. Since the majority of participants were start-ups, this report refers to all participants as such. Participants may also, *inter alia*, be referred to as firms, businesses and (successful) applicants depending on context.

We expect the beneficial effects of easier access to data to be visible in the performance of the participating start-ups in terms of investment and revenue achieved, as well as the number of jobs created.

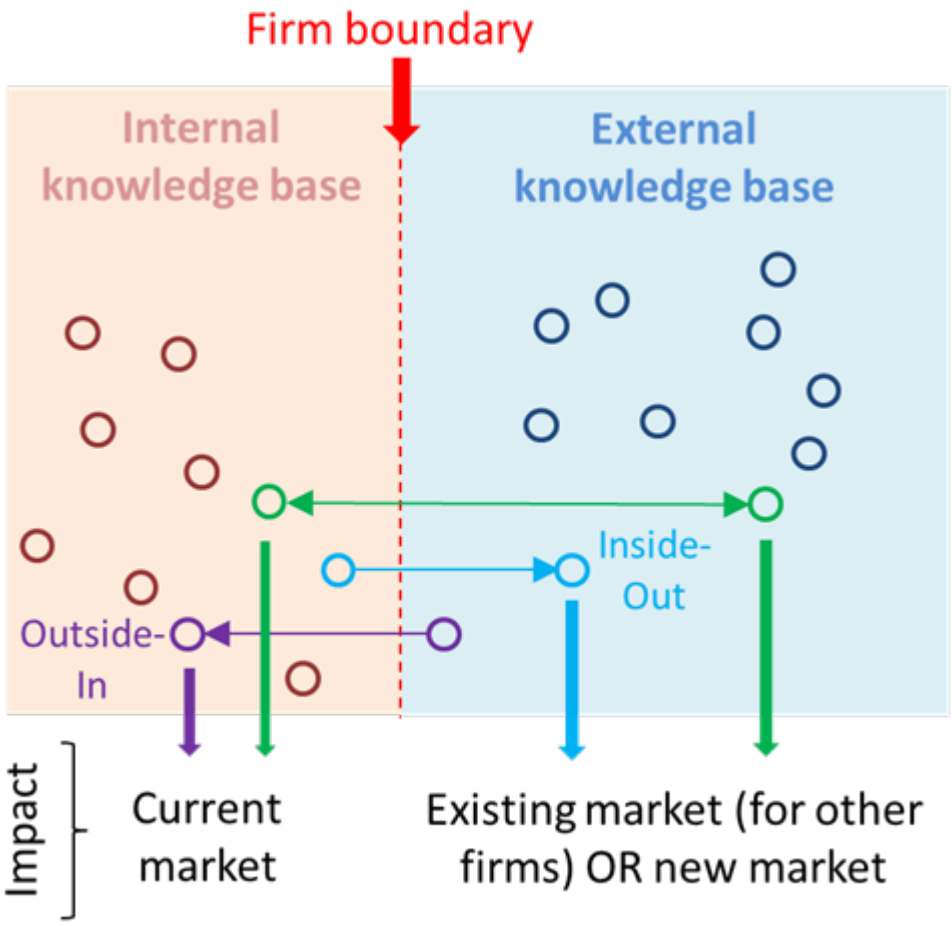
Measuring these effects quantitatively is one of the key objectives of this evaluation. In addition, qualitative assessment is required to ascertain whether the pilots delivered on expectations, especially in relation to innovation and the use of data; and what potential future impact from programme results on the data economy can be expected. The evaluation also addresses the scalability of the solutions that were developed by participant organisations; provides projections of impacts over a 3-year time horizon; and discusses data providers' willingness to work with innovators in the future.

1.2 The open innovation framework

1.2.1 Data Pitch as an implementation of open innovation

"Open innovation" is built around the insight that the inputs into the innovation process – such as data – are widely distributed across economic actors. More specifically, open innovation refers to organisations making greater use of external ideas and technologies to achieve their own purposes, and letting unused internal ideas and resources "go outside" for others to use (Chesbrough & Bogers, 2014).

Figure 1 The Open Innovation model



Source: London Economics based on Chesbrough & Bogers (2014), figure 1.5. p. 18.

The crucial insight is that there is both outside-in movement of knowledge and solutions (benefiting the data provider), but also inside-out movement (benefiting the start-up), and coupled innovation, resulting in close partnerships between data providers and start-ups.

- **Outside-In:** involves opening up a company's own innovation processes to external inputs and contributions, for example through acquiring or sourcing
- **Inside-Out:** requires organisations to allow unused and underutilised ideas and assets to go outside the organization for others to use in their businesses and business models
- **Coupled:** involves combining inflows and outflows of knowledge to collaboratively develop and/or commercialize an innovation (Chesbrough & Bogers, 2014).

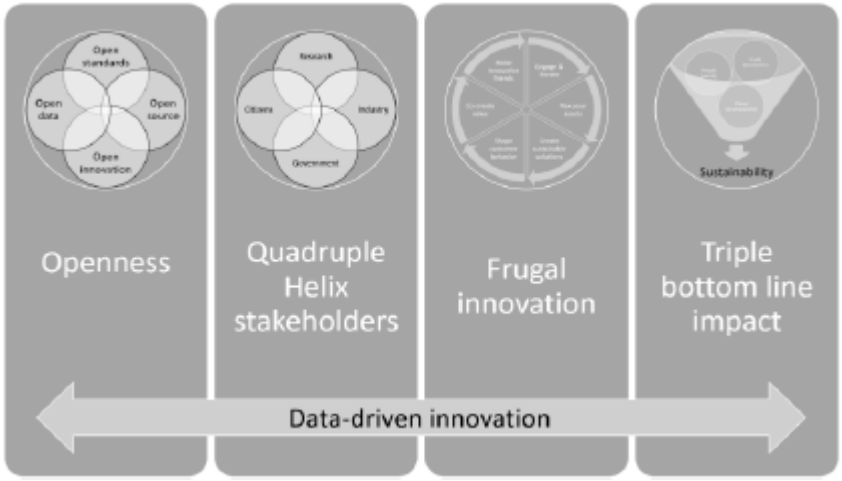
Data Pitch is a pioneering programme aimed at facilitating open innovation through data-sharing. Coupled open innovation is the most characteristic mode for Data Pitch, which helped data providers to develop their challenges, and facilitated the matching between data providers and start-ups. Data Pitch establishes a transnational data innovation ecosystem that creates collaboration between data providers on the one hand, and start-ups with fresh ideas for data-driven products and services on the other.

Data Pitch creates a "cross-sectoral, secure data experimentation facility", giving participants a "purposeful environment to test and nurture their ideas", while also providing the functions of a traditional business accelerator, supporting participating businesses through funding, technical, legal, marketing, and commercial assistance.

In addition, Data Pitch sets itself up as "an ecosystem enabler and digital innovation catalyst throughout the EU" beyond the acceleration phase that ended in November 2019. This means that Data Pitch participation creates know-how in the participating data providers that enables them to pursue open innovation, on their own or through other collaborative programmes, while demonstrating its usefulness through validated examples in the form of successful solutions appearing in the market.

The Data Pitch approach is rooted in a commitment to openness in its various manifestations (software, partnerships, data, etc.) and to the Quadruple Helix model of broad stakeholder engagement. This model acknowledges the value of interactions among citizens, government, industry, and academia, and aims to create impacts beyond immediate corporate revenues to create value for profit, people, and the planet.

Figure 2 The Data Pitch approach to open innovation



Source: Data Pitch grant proposal

1.2.2 Measuring innovation in Data Pitch

"An innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations." (OECD, 2005)

Innovation in Data Pitch comes from combining datasets, skills and knowledge in new ways, transferring existing products and services to new industries, and building new data-driven business models, rather than the invention of fundamentally new methods or first-of-a-kind products[3].

[3] Interviews confirmed that the start-ups all move in competitive markets where there are – sometimes close – substitutes for their solutions

Innovation is therefore not measured as a discrete outcome. Instead, innovation is embedded in the process of selecting the participating start-ups based on their ability to create a solution that is a new or improved way of using data providers' data to solve specific challenges, and ultimately generate commercial success for the start-ups.

1.3 Study approach

The study explores and evaluates the impact of Data Pitch on the two participant cohorts[4]. The study uses a mix of quantitative and qualitative data obtained from programme documentation, published sources, and primary data collection.

The main quantitative data collection for this impact assessment was a survey of the 47 start-ups funded through Data Pitch (successful applicants)[5]. 41 observations were collected. This survey was complemented with a survey of applicants who were not successful (9 observations)[6].

Qualitative information was collected through interviews with start-ups and data providers during September-November 2019:

- **17 interviews with start-ups[7] conducted over the phone and face to face, covering**
  - Background/motivation for participation
  - Data received
  - Use of funds/other support received
  - The solution
  - Benefits for the start-up
  - Reflections on the programme
- **6 phone interviews with data providers[8], covering:**
  - Background/motivation for participation
  - Data provided
  - Other support provided
  - Benefits for the data provider
  - Reflections on the programme

We also held consultations with Data Pitch staff (ODI, University of Southampton), in particular on the background and objectives of the programme, on the application and selection process, and on the support provided to participants.

Additional information was obtained from programme documentation provided by the ODI, including participants' applications and evaluation scores, and bi-weekly, 6-months & 12-months progress forms. External data sources that were used in the analysis include:

[4] See Annex 1.

[5] The questionnaire of the participants' survey is reproduced in Annex A3.4.

[6] The questionnaire of the unsuccessful applicants' survey is reproduced in Annex A3.5.

[7] Topic guide in Annex A3.2

[8] Topic guide in Annex A3.3

- Crunchbase
- Orbis

A focused literature review was conducted about the impact and performance of technology accelerators and incubators[9].

Six case studies illustrate the achievements of selected start-ups in the provider and sectoral challenges. The case studies cover details on solution and data used, as well as the role of the data providers and business outcomes.

The data collected through the surveys, together with performance data collected by the Data Pitch consortium formed the basis of a counterfactual analysis of programme impacts (Section 3.8) and a simulation-based estimation of future benefits (Section 3.9). A methodology for a quantitative (econometric) evaluation of the programme is documented in Annex 4.

[9] A key reference is Bone et al. (2019).

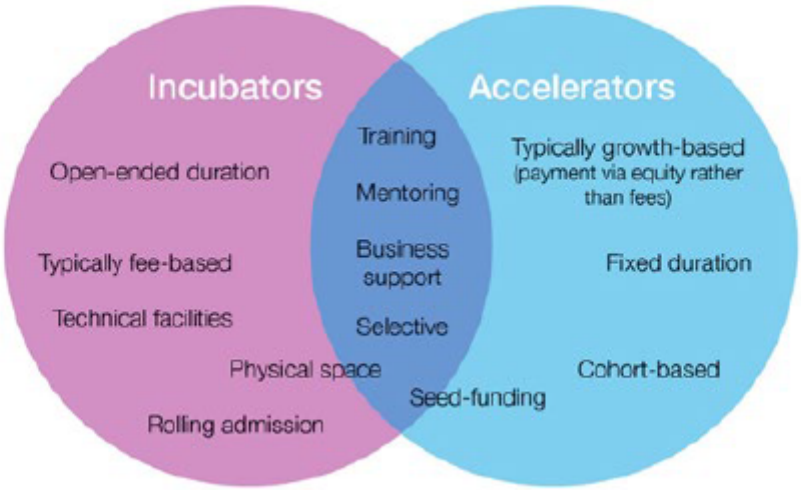
2 The Data Pitch open innovation programme

2.1 How the Data Pitch model fits in the start-up support landscape

Data Pitch shares many characteristics with incubators and, even more so, accelerators. Data Pitch is distinguishable from other accelerators through its focus on open innovation and data.

Data Pitch shares many of the characteristics of incubators and accelerators, with the cohort-based approach and fixed duration placing it more in the accelerator spectrum.

Figure 3 Defining characteristics of incubators and accelerators



Source: Bone et al. (2019), Figure 1

According to Miller & Bound (2011), an accelerator programme has five main features:

- Application process open to all, yet highly competitive.
- Provision of pre-seed investment, usually in exchange for equity.
- A focus on small teams not individual founders.
- Time-limited support comprising programmed events and intensive mentoring.
- Cohorts or 'classes' of start-ups rather than individual companies.

Like other accelerators (and unlike incubators), Data Pitch offered its services through an intensive cohort-based programme of limited duration (6 months, in line with the 3-12 months range usually seen for accelerators) after an open competition to enter the programme[10]. Like other accelerators, Data Pitch also focused on services over physical space. However, unlike the majority of accelerators, Data Pitch's direct funding of participating start-ups does not take the form of equity investment. Another distinguishing feature of Data Pitch is its virtual setup.

[10] See Bone et al. (2017), Clarysse et al. (2015)

Accelerator programmes are set up to benefit start-ups primarily, but often also have wider economic development objectives. Miller & Bound (2011) lists the beneficiaries of accelerators as follows:

- Angel investors
  - Reduce the need for due diligence as that role is performed by accelerator.
  - Reduce the cost and time required to find new companies to work with.
  - Ability to meet other investors and company founders.
- Venture Capital Firms
  - Improve deal pipeline, creating more high-quality start-ups.
  - Get first sight of new technology and ability to map trends in start-ups.
  - Ability to meet other investors and company founders.
- Large Tech firms
  - Talent scouting for new employees.
  - New customers for their platforms and services.
  - Associate their brand with supporting new businesses.
- Other start-up founders
  - Talent scouting for new employees.
  - Rapidly create a very high-quality business network.
  - Meet customers and later-stage investors that might be relevant to their businesses.
- Service providers (e.g. accountancy firms, law firms, PR firms)
  - New customers in the form of the start-ups the accelerators support.

While Data Pitch in its accelerator role is likely to produce the same benefits, it is distinguishable from other accelerators through its focus on open innovation, which shaped the programme design as well as the recruitment mechanism. Furthermore, Data Pitch also contributed to the wider data innovation ecosystem by providing a framework and proven case studies of open innovation which may inspire other data holders to engage with open innovation.

The following section provides a description of the Data Pitch programme.

2.2 Description of the programme

Data Pitch is more than an accelerator. It combines a thorough preparation phase to identify suitable challenges with technical and business support, networking opportunities, and funding, to achieve impact through open innovation.

Data Pitch is an EU-funded programme with the aim of fostering open innovation for start-ups and SMEs that work with data.[11] The programme consists of multiple components which all help to facilitate this acceleration:[12]

- A grant (equity-free investment) of up to €100,000;
- introductions to investors;

[11] https://datapitch.eu/wp-content/uploads/2018/07/Guide-for-applicants-call-2018vfq.pdf

[12] https://datapitch.eu/about-us/

- 6-month business accelerator programme with ODI and Beta-I;
- peer-networking and support via meetups;
- access to training materials and webinars by Data Pitch experts;
- legal counsel for IP protection;
- drafting of a data sharing agreement between Data Pitch partners.

The key distinguishing feature of Data Pitch is that it matched data providers that possessed data that they wanted to use for open innovation, with start-ups with the wherewithal to exploit this data. Start-ups in the "provider" challenges were matched by the Data Pitch consortium, whereas start-ups in the "sector" or "open" challenges had to find their own data provider. All participants had to have a partnered data provider.

Data Pitch aimed for the programme to involve less arduous paperwork (akin to the standard accelerator process) than a usual EU funding competition[13] with the application form only requiring details on basic company information, details about their team and their idea proposal.

2.3 Pre-selection

The challenges for the Data Pitch calls were chosen based on extensive consultation with data providers, industry experts and scientific communities. The result was a collection of challenges created with the aim of achieving the highest levels of impact in sectors identified for their potential for impact, access to data and opportunity for additional network effects.

Prior to the accelerator period, Data Pitch conducted consultations with key stakeholders (such as data providers, industry experts and scientific communities) to identify the most effective domains for these challenges to address. The challenges (both data provider and sectoral) were created with the following concepts and ideas[14]:

- Be inclusive and engaged – provide a context and forum for fostering discussion and collaboration;
- Target market failures;
- Have clear requirements – ensure results are measurable;
- Be solution agnostic – to allow for participants to organically develop a solution which is not constrained by overly prescriptive, technical and non-technical requirements.

For Call 1, the final challenges were within the sectors of[15]:

- Retail (data provider: Sonae);
- Data Analytics (data provider: Spazio Dati);
- Sports & Recreation (data provider: imin);
- Transport (data provider: Deutsche Bahn);
- Data Management (data provider: Uniserv);
- Health & Wellness;

[13] Correspondence with the Open Data Institute.

[14] Data Pitch deliverable D3.4: First Data Pitch Consultations

[15] https://datapitch.eu/challenges/challenges2017-2/

- Empowering Users Online;
- Lifelong Learning;
- Living;
- Smart Manufacturing;
- Tourism;
- Open innovation (open challenge).

These sectors were identified by Data Pitch as areas in which solutions could provide high impact; sectors which also contained closed datasets which were not restricted by tight regulation as well as sectors with opportunity to leverage additional network effects (by targeting tracks which are a priority for local and international governments). These overlap with areas which the Big Data Value Association Public-Private-Partnership have also identified as areas in which data innovation can provide significant gains.[16]

These challenges were created as a result of consultations carried out by Data Pitch. The components of these consultations included a survey, a workshop and interviews with identified data providers. From these interviews, specific challenges were created in conjunction with the participating data providers.

For Call 2, Data Pitch identified the BDVA areas which were either not covered in Call 1, or were only partially addressed. This led to an increase in the number of challenges in additional sectors. The final challenges were within the domains of:[17]

- Pharmaceuticals;
- Automotive;
- Energy;
- Finance;
- Telecoms;
- Privacy & Consent Control;
- Smart Transport;
- Personalised Entertainment (data provider: Altice Labs);
- Text Mining & Analytics (data provider: Bloomberg);
- Smart Manufacturing (data provider: Greiner International Packaging);
- Sustainable Food Supply Chain (data provider: GROW);
- Customer Needs Prediction (data provider: Konica Minolta);
- Healthcare (data provider: José de Mello Saúde);
- Multimodal Transport (data provider: MASAI);
- Weather and Climate Change (data provider: MET Office);
- Open Challenge.

[16] The key verticals of the Big Data Value Association are: Environmental and geospatial data; Energy; Mobility, transport and logistics; Manufacturing and production; Public sector; Healthcare; Media and Content; Finance; Telecoms; Retail; Tourism.

[17] https://datapitch.eu/challenges/challenges-2018/

An Open Challenge was also issued in Call 1 and 2 to capture any novel ideas that had not been pre-identified from this consultation and pre-loading phase.

2.4 The Data Pitch accelerator

The Data Pitch accelerator was a 6-month accelerator programme which provided successful applicants with access to funding, mentors and advisors. Project monitoring arrangements involved a combination of meetings with advisors, workplans and deliverables at set milestones.

The Data Pitch programme involved a 6-month accelerator programme delivered by Data Pitch partners Beta-i and the ODI (Open Data Institute). Beta-i also runs several open innovation programmes in several sectors, including smart cities, tourism, fintech and healthcare. One of these programmes includes the Lisbon Challenge, a Portuguese accelerator which has funded over 70 start-ups, which have collectively raised over €75 million.[18] Other examples of programmes ran include the Blue Tech Accelerator, The Journey & Free Electrons. The ODI had previous experience with the Open Data Incubator for Europe (ODINE) project; a virtual accelerator which funded 57 companies.

At the beginning of this accelerator, these start-ups created a workplan and budget for their projects. These workplans would serve as a guideline for each start-ups major milestones, deliverables and goals they would aim to achieve. These workplans, which were created in negotiation with Data Pitch and their data providers, were not expected to be followed strictly; Data Pitch acknowledged that unexpected challenges and opportunities could affect how closely these start-ups could abide by these initial plans.

During the programme, the start-ups had access to a mentor, whom they could meet with throughout the acceleration period for guidance and support; over 70% of these mentors involved in the first cohort have kept in touch with their allocated start-ups after the programme ended[19]. These mentors included academics at the University of Southampton, and other professionals specialised in technology-related topics such as data science and AI, as well as experts in business-related fields such as marketing, sales and consumer behaviour[20]. Meetings with mentors were arranged by the companies and mentors themselves on an ad-hoc basis[21].

Alongside these mentors, each start-up was also assigned an advisor. These advisors were the company's first point of contact for concerns, queries and general comments. The aim of the advisors was to provide support, guidance and help with setting interim goals during the acceleration period. Each company met with their advisor roughly every two weeks[22]. Additionally, a monthly group session would be arranged in which an advisor and all their matched companies would meet to identify shared problems and solutions.

At these bi-weekly meetings, these companies agreed with their mentor on next steps and current progress. Furthermore, bi-weekly progress was tracked by the Data Pitch consortium.

[18] http://www.lisbon-challenge.com/#program

[19] https://datapitch.eu/news/make-a-difference-become-a-mentor/

[20] https://datapitch.eu/network/mentors/

[21] https://datapitch.eu/wp-content/uploads/2018/08/PUBLIC_DataPitch_.5.1-Data-Pitch_IncubationServices-1.pdf

[22] ibid.

In addition to the regular updates, the programme contained three milestone touchpoints. For each milestone, a number of activities and performance goals were agreed with the start-ups as part of the initial workplan formulation:[23]

- **Activities:** What the company will achieve in the milestone, e.g. 'Write the code for data pre-processing pipeline in Python and Scala'
- **Key performance indicators (KPIs):** A quantitative outcome to be achieved, e.g. 'Cold calls to 50 potential customers'
- **Deliverables:** Outputs and evidence of progress towards completing these KPIs, e.g. Business plans, interview transcripts, demos.

Payment of the grant was distributed over the acceleration phase as follows: 30% of the grant was paid at the start of the accelerator, 30% was paid after the 4-month milestone review, and 40% was paid after the final milestone had been met.[24]

2.5 Additional benefits & programme services

To help start-ups maximise their impact, Data Pitch provided additional benefits which were chosen with the aim of helping start-ups deal with business and technical challenges. These benefits involved software perks, workshops and networking events. The support package was tailored to the needs of each start-up.

Data Pitch provided several additional services. These services were defined and curated from the Data Pitch consortium's combined knowledge from ODINE, the multitude of innovation programmes ran by Beta-i and from the collective experience of the other consortium organisations. These services included:[25]

- **Workshops**
  - 10 workshops were held per cohort, either virtually or in person, and delivered training on key themes relevant for the Data Pitch cohorts; themes included, *inter alia*, marketing, design, market fit and EU General Data Protection Regulation (GDPR).
  - Workshops were delivered by representatives from each member of the Data Pitch Consortium, or by professionals from external organisations.
- **Founder stories**
  - Founder stories refer to a series of sessions with founders of established start-ups. These 'intimate private sessions' allowed Data Pitch start-ups to gain a deeper understanding and first-hand experience of starting and scaling-up a company in the technology industry.
- **Networking**
  - Start-ups had access to the start-up network of ODI and Beta-i. The aim of this networking was to both increase exposure and allow these start-ups to meet with key stakeholders, such as investors, clients and potential future partners.

[23] Correspondence with the Data Pitch consortium.

[24] https://datapitch.eu/wp-content/uploads/2018/08/PUBLIC_DataPitch_.5.1-Data-Pitch_IncubationServices-1.pdf

[25] ibid.

- Other networking opportunities shared with start-ups included various events that provided opportunities to pitch and to meet clients and investors.

- **Perks**
  - Provision of perks and discounts such as Amazon Web Hosting credits and access to tools such as Hubspot for Start-ups, Segment (analytics platform) & TestArmy (software testing tools). These perks were provided according to the needs of the companies.

2.6 Selection process

Over the two Data Pitch calls, 47 start-ups, out of a total number of 239 applicants, were successful in gaining admission to the programme. 22 of the different challenges were matched with start-ups, with one data provider challenge in each call left unmatched. 18 start-ups were selected for funding in Call 1 and this increased to 27 in Call 2.

There were two Data Pitch calls. The first call was issued on the 1st July 2017 and the second a year later, on the 2nd July 2018. Both calls used the F6S platform to receive and manage applications from start-ups.

Applicants had to be individual companies (no consortia) and qualify as SMEs at the time of submission[26]. Additionally, they also had to be established and working in an EU member state or a non-EU Horizon 2020 partner country.

Applicants had to submit a short proposal (around 5 pages), supporting documents and basic information about the company. Additionally, applicants also had to provide a 12-slide pitch deck and a 1-minute video explaining why Data Pitch should fund their teams.[27]

The calls took the format of a list of challenges[28]. Applicants were invited to an interview based on the strengths of their applications. Three experts (a sector, business and tech expert) were present at these interviews. Furthermore, data providers were also present to judge 'their' challenge. These experts questioned the applicants over their proposals and returned either a Yes or No verdict.

Table 2 Outcome of Judge Verdicts

Judge Verdicts	Outcome
Yes 3 times	Funding granted
Yes 2 times or No 2 times	Application reviewed further
No 3 times	Funding rejected

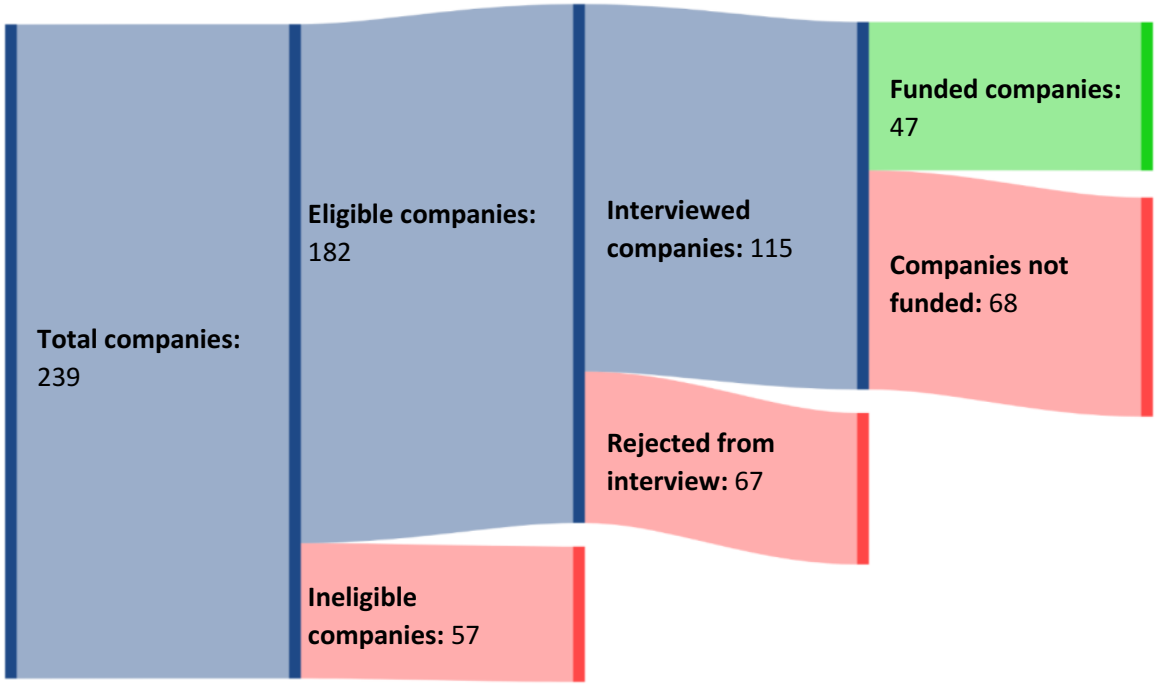
The figure below shows the flow of applicants through the selection stages for both cohorts.

[26] To qualify as an SME under the EC definition, a firm must have a staff headcount of less than 250 and a turnover of less than or equal to 50 million euros. See https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en.

[27] D4.3 Summary of Round 2 ODI

[28] See Annex 1.

Figure 4 Selection of Data Pitch participants



Note: The data in this diagram refers to both cohort 1 and 2 companies

Source: London Economics based on data provided by Data Pitch

2.6.1 Cohort 1: process & outcomes

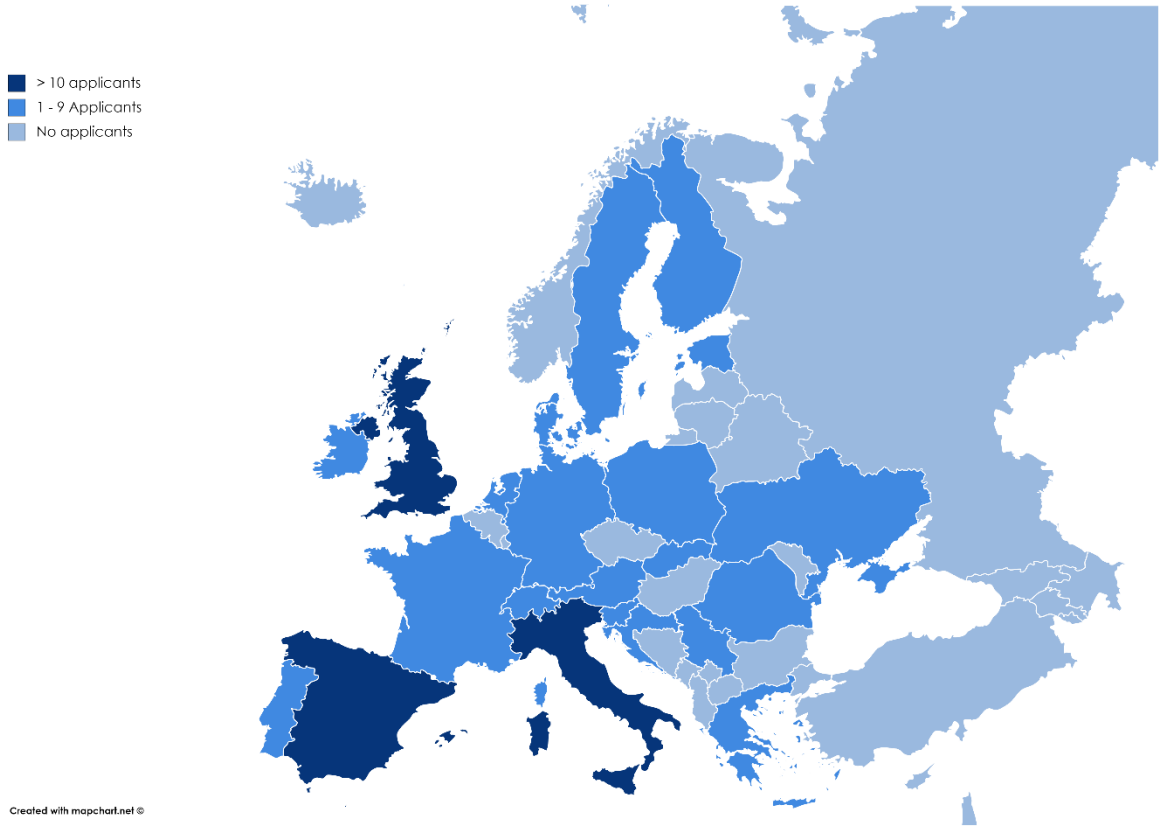
Data Pitch's publication strategy was to inform 'local technology hubs' of the competition, who would then disseminate this information to start-ups and SMEs in the area.[29] These 'local hubs' include organisations such as accelerators, incubators and community spaces. Additionally, Data Pitch scouted for applicants at several events which took place in London, Lisbon, Berlin, Stockholm, Copenhagen, Cologne & Munich.

The first call contained 12 challenges; 5 data provider challenges (with a data provider arranged by Data Pitch), 6 sector challenges (requiring applicants to find a data provider) and 1 'open' challenge allowing participants to provide a proposal without sector or specific challenge restrictions.

142 applications were received, 112 of which were eligible for funding. These 112 applications were scored to determine the final set of 57 applicants who were invited to interview. The geographical spread of the applicants can be seen in Figure 1.

[29] Correspondence with the Open Data Institute.

Figure 5 Map of Data Pitch Call 1 Applicants



Note: Data Pitch infographic recreated by LE using mapchart.net

Source: Data Pitch deliverable D.4.2.: Summary of Round 1

Applications were judged using three criteria:

- strength and novelty of the idea (idea);
- value proposition and potential scale (impact); and
- the team and budget (team).

These three categories were scored out of 30, 30 and 40 respectively. Each application was reviewed independently by two reviewers. If the scores for an application differed drastically between reviewers, a review board would discuss and resolve discrepancies.[30] An application had to score over 60 to be invited to interview.[31] These interviews took place in London and lasted 45 minutes. Outcomes of the interviews determined whether a start-up received funding.

18 start-ups received funding; 13 completed sectoral challenges and 5 completed a data provider challenge. Start-ups were match to challenges as follows:[32]

- Health & wellness [sector challenge] (4)
- Smart manufacturing [sector challenge] (4)

[30] Correspondence with the Open Data Institute.

[31] Data Pitch deliverable D.4.2.: Summary of Round 1

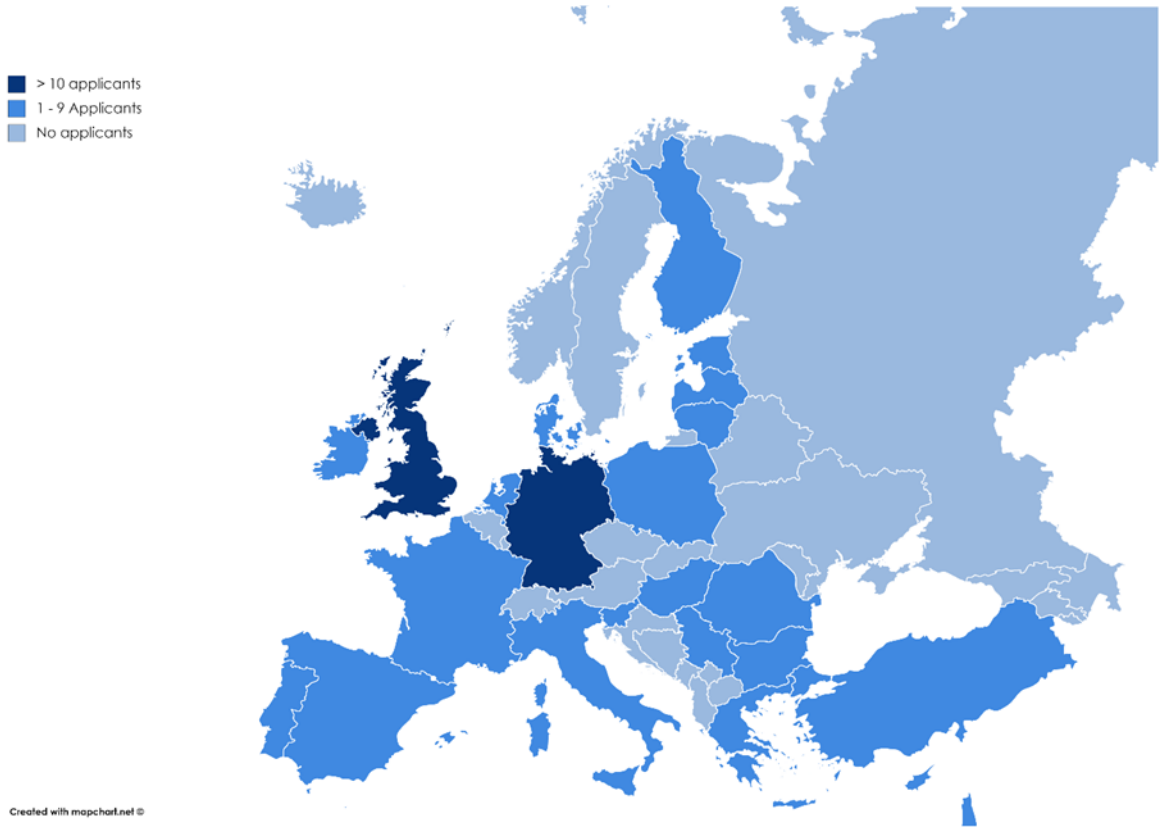
[32] Ibid.

- Tourism [sector challenge] (3)
- Data management [data provider challenge; data provider: Uniserv] (2)
- Data analytics [data provider challenge; data provider: SpazioDati] (2)
- Transport [data provider challenge; data provider: Deutsche Bahn] (1)
- Retail [data provider challenge; data provider: Sonae] (1)
- Empowering users online [sector challenge] (1)

2.6.2 Cohort 2: process & outcomes

Call 2 was similar to the first but increased the number of challenges to 16 (8 data provider challenges, 7 sector challenges and 1 open challenge). This call focused on reaching countries which were under-represented in call 1. This meant that Data Pitch scouted at events in countries such as Poland, Norway and Malta (alongside events in Portugal, Germany, The Netherlands and the UK). The geographical spread of applicants can be seen in Figure 6.

Figure 6 Map of Data Pitch Call 2 Applicants



Note: Data Pitch infographic recreated by LE using mapchart.net.

Source: Data pitch deliverable D.4.3.: Summary of Round 2

97 applications were received and 70 were eligible. These 70 were reviewed by a team of 8 reviewers (each application being reviewed twice) to arrive at a final list of 58 who were invited for an interview. The same criteria used in call 1 (idea, impact & team) was also used in call 2.

29 start-ups were selected for funding, and matched as follows:[33]

- Smart manufacturing [data provider challenge; data provider: Greiner] (5)
- Customer needs prediction [data provider challenge; data provider: Konica] (3)
- Open innovation challenge (3)
- Energy [sector challenge] (3)
- Personalised entertainment [data provider challenge; data provider: Altice Labs] (2)
- Weather and climate change [data provider challenge; data provider MET office] (2)
- Finance [sector challenge] (2)
- Privacy & consent control [sector challenge] (2)
- Smart transport [sector challenge] (2)
- Text mining & analytics [data provider challenge; data provider: Bloomberg] (1)
- Sustainable food supply chain [data provider challenge; data provider: GROW] (1)
- Healthcare [data provider challenge; data provider: José de Mello Saúde] (1)
- Pharmaceuticals [sector challenge] (1)
- Automotive [sector challenge] (1)

For call 2, a greater number of applicants received funding. Additionally, four applicants applying to call 2 previously applied to call 1; two of these four companies were successful in receiving funding in call 2. This success could be due to new challenges which were more aligned with these companies. This success may also be attributed to possible learning effects occurring between the two calls, or due to the changes in the pre-selection phase between call 1 and 2.

In addition to the wider range of challenges for cohort 2, call 2 also allowed applicants to be interviewed remotely. This ensured that under-resourced, or newly established, start-ups could ensure that both their tech and management people were at the interviews. This would allow for interview questions to be answered in more depth, as well as allow Data Pitch to gain a better understanding of applicants when making their decisions.

[33] Data pitch deliverable D.4.3.: Summary of Round 2

3 Programme performance

3.1 Performance objectives

A successful programme will show evidence of sustainable growth and investment, increase in both sectoral and geographical ecosystems, and generation of data-driven use cases.

The Data Pitch consortium has defined three overall indicators along which to measure the success of the Data Pitch project:[34]

- *"A sizable number of data-driven businesses are established, and several lighthouse start-up ideas receive wide public attention";*
- *"There is an expanding ecosystem of organisations collaborating to offer technology and methodological support for data innovation labs, across sectors and borders"*; and,
- *"The concept of data-driven business, including social entrepreneurship, is well developed and receives substantial investment in enterprise ecosystems, investor circles, and other funds."*

Therefore, a successful programme will have evidence of:

- establishment and expansion of companies which have sustainable growth and are receiving substantial investments;
- an increase in both sectoral and geographical ecosystems and evidence of cross-border/sector collaboration; and,
- impact generated by data-driven businesses and big-data use cases.

3.2 Sectors of activity

Many of the participating start-ups worked across sectors, with the aim to enter new sectors being a key motivator for participating in Data Pitch. A concentration of participating start-ups can be found in the health, financial and transport/mobility sectors.

The participants cover a wide variety of activities and many operate across sectors. Moreover, as young companies, many have not yet reached a stage where it is possible to clearly identify core activities or sectors of operation. In fact, entering new sectors has been a key motivator for participating in Data Pitch for a number of start-ups. As a result, this report does not attempt to assign start-ups to standard industry classifications. Instead, we use self-reported target markets as the most useful sectoral indicator.

Table 3 Start-ups primary customer for Data Pitch-enabled solution

Clinics	Utilities and owners of large assets
Hospitals	Offshore wind and energy sector
Medical staff	Heavy industry
Emergency services and local authorities	Flight companies and mountain stations
Local authorities	Building managers
Transportation and local authorities	Retailers

[34] Data Pitch Grant Agreement. Shared with London Economics by the consortium.

Highway managers	Online retailers
Vehicle fleet managers and automotive industry	Food service industry
Car manufacturers and insurers	Food companies
Bank and insurance companies	Tourism
Financial institutions	Customer-facing businesses
Bank and other lenders	Business analysts
Banks and regulators	Data scientists and engineers
Finance, insurance and manufacturing	Pay TV/Cable companies
Investment banks/large companies	Personal data managers and consumers
Large companies	Consumers
Businesses with high customer volume	Operators of industrial machinery

Source: Successful applicants survey

The health, financial and transport/mobility sectors stand out as being the focus of several start-ups. However, many of the solutions are inherently cross-sector. Overall, the focus is mostly B2B rather than B2C. This reflects the fact that many of the challenges were formulated for the direct benefit of the data providers, which makes for an inherent B2B focus.

3.3 Geographical distribution

Data Pitch reached start-ups based in 13 countries, with cohort 2 broadening the reach of the programme. The United Kingdom represented a large proportion of all participants. The UK, together with Spain, also provided the highest number of participants invited for interviews.

Start-ups in cohort 1 were based across 10 countries, whereas start-ups in cohort 2 were based across 13 countries. Call 2 reached additional countries in Eastern Europe (such as Latvia and Serbia).

UK represents 26% (cohort 1) and 20% (cohort 2) of funded companies. The decrease in the proportion suggests that the attempt to increase representation across the eligible countries was successful.[35]

Table 4 Geographical spread of funded companies by cohort

Country	Number of funded applicants in cohort 1	Number of funded applicants in cohort 2
UK	5	7
Germany	3	6
Ireland	0	3
France	2	2
Spain	2	2
Greece	1	2
Portugal	1	1
Denmark	2	1
Netherlands	1	1
Italy	1	0
Romania	0	2
Serbia	0	1

[35] Correspondence with the Open Data Institute.

Country	Number of funded applicants in cohort 1	Number of funded applicants in cohort 2
Latvia	0	1
Total	18	29

Source: Data Pitch[36]

Table 5 shows the geographical spread of successful and unsuccessful applicants. Notable points are:

- Applicants from **Italy** had an 8% success rate of achieving funding. Only 1 out of the 12 interviewed were successful.
- 13 countries were unsuccessful at obtaining funding.
- The **UK** and **Spain** had the highest number of interviews but relatively low success rates at 27% and 17% respectively.
- **Germany**, with the third highest number of interviews at 15, had the third highest success rate at 60%. This resulted in 9, or just under 20%, of the total funded companies being based in Germany.

Table 5 Geography of successful and unsuccessful applicants

Country	Unsuccessful funding (total)	Unsuccessful after interview	Successful funding	Proportion of total applicants successful	Total applicants
AT	3	2	0	0%	3
BE	1	0	0	0%	1
CH	1	0	0	0%	1
DE	6	5	9	60%	15
DK	2	1	3	60%	5
EL	2	1	3	60%	5
ES	20	10	4	17%	24
FI	2	0	0	0%	2
FR	6	3	4	40%	10
GR	1	1	0	0%	1
HR	1	1	0	0%	1
HU	2	2	0	0%	2
IE	2	0	3	60%	5
IL	1	1	0	0%	1
IT	11	8	1	8%	12
LT	1	1	0	0%	1
LV	0	0	1	100%	1
NL	1	1	2	67%	3
PL	1	1	0	0%	1
PT	6	4	2	25%	8
RO	0	0	2	100%	2
RS	1	1	1	50%	2
SE	3	1	0	0%	3

[36] These figures can be accessed here: https://datapitch.eu/data-pitch-start-ups/

Country	Unsuccessful funding (total)	Unsuccessful after interview	Successful funding	Proportion of total applicants successful	Total applicants
SK	2	1	0	0%	2
SL	2	0	0	0%	2
UA	1	1	0	0%	1
UK	33	19	12	27%	44
N/I[a]	2	2	0	0%	2
Total	114	67	47	29%	161

Note: Number of interviewed candidates is slightly lower than Data Pitch reported number due to some candidates interviewing in both cohorts. [a] Country could not be identified.
Source: London Economics analysis

3.4 European collaboration

24 out of 47 (51%) start-ups were partnered with a data provider located in a different country. In the data provider challenge, all but one of the 18 partnerships (94%) was a European collaboration.

An objective of Data Pitch is to foster cooperation between firms from different parts of Europe. As part of the accelerator programme, some start-ups were partnered with data providers from other Member States. The table below shows the extent of cross-border matching within Data Pitch.

Table 6 Countries of start-ups matched with data providers

Data provider	Country of DP	DE	EL	ES	LV	NL	PT	RO	RS	UK	Total
Altice Labs	PT	0	0	0	0	1	0	0	0	1	2
Bloomberg	US	0	0	0	1	0	0	0	0	0	1
Deutsch Bahn	DE	0	0	0	0	0	1	0	0	0	1
Greiner Packaging	AT	1	1	1	0	0	0	0	0	1	4
Grow Observatory	UK	0	1	0	0	0	0	0	0	0	1
José de Mello Saúde	PT	0	0	0	0	0	0	1	0	0	1
Konica Minolta	JP	1	0	0	0	0	0	1	1	0	3
Met Office	UK	0	0	1	0	0	0	0	0	1	2
Sonae	PT	0	0	0	0	0	0	0	0	1	1
Uniserv	DE	0	0	1	0	1	0	0	0	0	2
Total		2	2	3	1	2	1	2	1	4	18

Note: Red numbers indicate domestic partnerships.
Source: London Economics analysis

In the data provider challenges, only 1 (out of 18) start-up was matched with a provider operating in the same country (Met Office). This meant that **17 start-ups** were matched across European countries as part of Data Pitch. Data Pitch succeeded in facilitating international collaboration across Europe.

Table 7 Countries of start-ups in the sector and open challenges matched with data providers

Country of DP	CH	DE	DK	EL	ES	FR	IE	IT	UK	Total
CH	0	1	0	0	0	0	0	0	0	1

Country of DP	CH	DE	DK	EL	ES	FR	IE	IT	UK	Total
DE	0	4	0	0	0	0	0	0	0	4
DK	0	0	1	0	0	0	0	0	0	1
EL	0	0	0	1	0	0	1	0	0	2
ES	0	0	0	0	0	0	0	1	0	1
FR	0	1	0	0	0	3	0	0	0	4
IE	0	0	1	0	0	0	1	0	1	2
IT	0	0	0	0	0	0	0	0	0	0
UK	0	0	1	0	0	0	0	0	4	5
Total	0	6	3	1	0	3	2	1	5	21

Note: Data provider details for three start-ups could not be located, an additional start-up did not have a partnered data provider. Red numbers indicate domestic partnerships.

Source: London Economics Analysis

For start-ups in the non-data provider challenges, 14 entered Data Pitch with a partnered data provider which were located in the same country; **7 start-ups** were matched across international borders. Between the data provider, sector and open challenges, **24 start-ups** were matched with a non-local data provider.

3.5 Start-up performance

During the programme, start-ups on average increased sales by €36,554, received investment of €82,448 and increased employment by 2. Cohort 1 generated, on average, sales of €134,379, investment of €71,334 and additional employment of 1 in the 6 months following the programme. By the end of the programme, the Return on Investment was already 23%, increasing to 91% 6 months after the end of the programme.

Throughout the programme, Data Pitch asked start-ups to update their progress using the following metrics:

- total amount of sales;
- investments;
- realised efficiencies;[37]
- change in employment; and,
- whether the applicant is looking for further funding.

Data Pitch required start-ups to record these metrics in bi-weekly progress reports over the six-month accelerator period. Table 8 presents a high-level summary of these different metrics.

Table 8 Summary of start-up progress during the Data Pitch programme

Indicator	Observations	Mean	Minimum	Maximum	Total
Sales	47	38,682	0	420,000	1,818,045
Investments	47	82,448	0	1,600,000	3,875,060
Efficiencies	47	17,168	0	318,000	806,890
Employment	47	2	-1	12	114

[37] Efficiencies realised refers to a loose metric designed to capture time saving that the Data Pitch programme allowed a start-up to pass on to their clients. This monetary value does not refer to monetary value saved for the start-up, but instead for their clients/customers.

Note: Values are in euros (other than Employment). Investments include Grants, BAs, VCs etc. Bi-weekly monitoring updates cover approximately 24 weeks.

Source: Bi-weekly monitoring updates

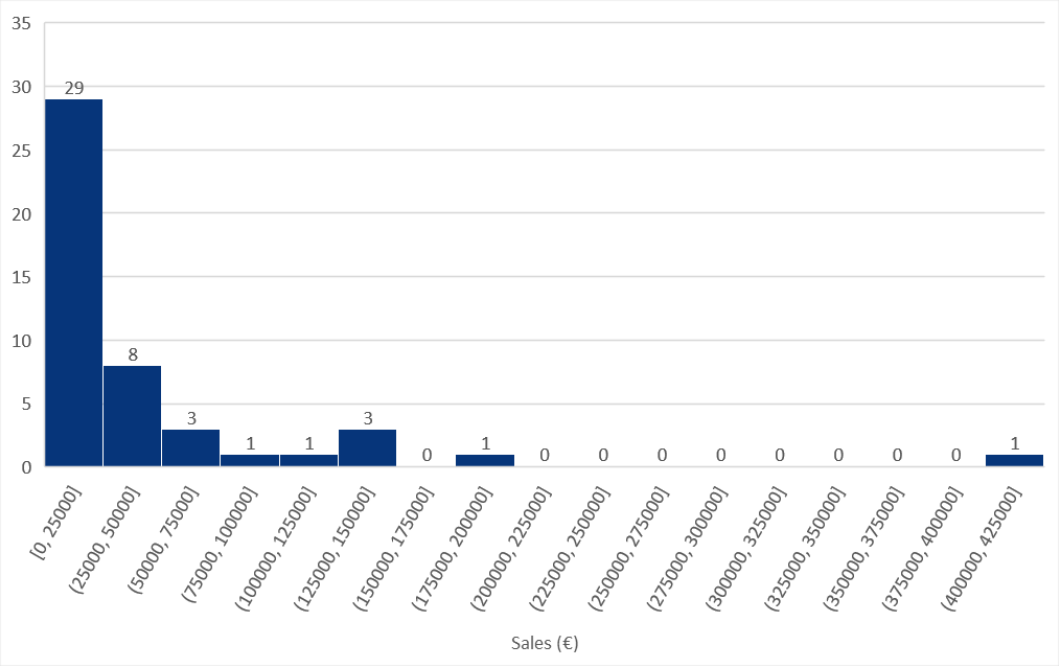
During the 6-month acceleration phase of the Data Pitch programme, on average across both cohorts, start-ups increased their sales by €36,554, received investments of €82,448, realised efficiencies of €17,168 and increased their employment by 2 employees.

Across the 47 start-ups, 21 (45%) reported that they did not raise any revenue from sales; median sales amount (including firms with zero sales) was €1,550. 75% of all firms reported sales of €43,680 or lower. The highest three sales revenues raised were €420,000, €193,950 & €145,625.

The top three sales figures across both cohorts were obtained by companies which had been established for around four years prior to their participation in Data Pitch. Conversely, a proportion of start-ups which raised zero in sales were young (having been established for Data Pitch, or very soon before). This suggests age and maturity level of the firm may have influenced revenue generation during this period.

Four of the five highest revenues were obtained by start-ups who were completing a data provider challenge (with four of the top ten having completed a sectoral challenge). Of the 21 firms which reported zero in revenue, 8 were completing a data provider challenge, with the remaining 13 competing either a sectoral or open challenge. Therefore, 34% of start-ups completing data provider challenges and over half (54%) of start-ups completing sectoral or open challenges had raised zero in sales revenue.

Figure 7 Distribution of sales for Data Pitch participants during the accelerator period



Note: Y-axis refers to number of start-ups, total N = 47.

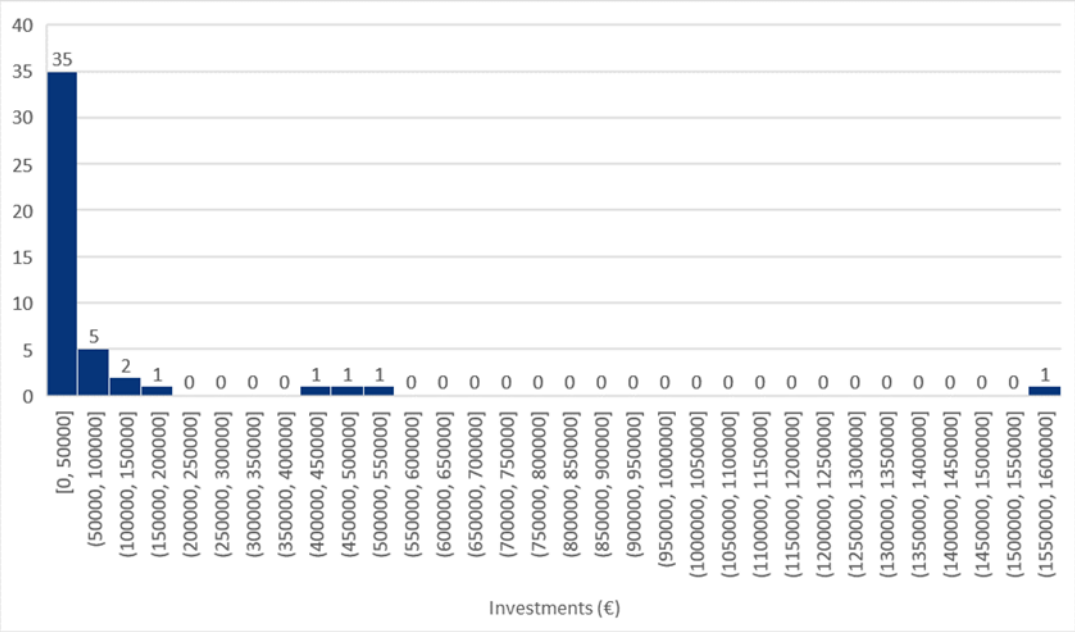
Source: Bi-weekly monitoring updates

Twenty (43%) start-ups raised a non-zero amount of investment (other than the funding received from Data Pitch). 8 start-ups raised investment up to €45,000 and the remaining 12 firms raised over €45,000. The highest three levels amounted to €1,600,000, €507,000 and €460,000. All three of these start-ups were completing sector challenges, and two of these firms had raised zero in sales

during the acceleration period. The start-ups which raised the top two investment figures were established within a year of entering Data Pitch (with the third established within two).

For the firms which raised zero in investments, just over half (14 out of 27) were completing Data Provider challenges. Just under 61% of all start-ups completing data provider challenges had raised zero in additional investments; for comparison, 50% of start-ups completing sectoral and open challenges did not receive any additional investments.

Figure 8 Distribution of investments for Data Pitch participants during the accelerator period

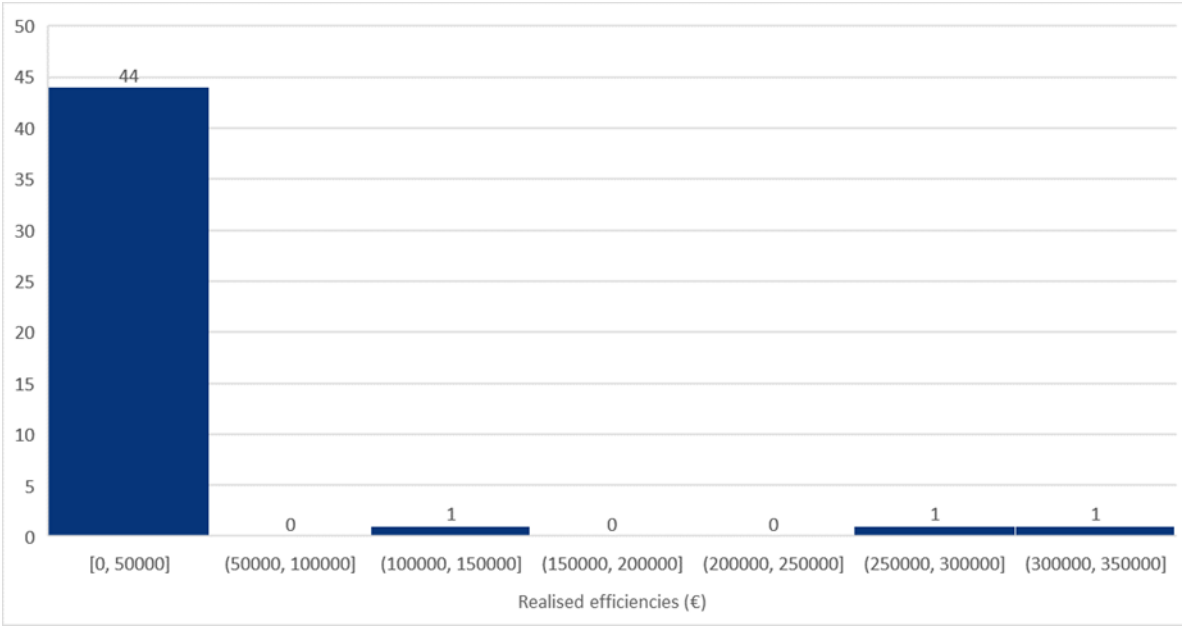


Note: Y-axis refers to number of start-ups, total N = 47.

Source: Bi-weekly monitoring updates

11 (23%) start-ups realised efficiencies, saving a positive amount of money (in timesaving) for their clients/customers. The value of these efficiencies ranged from €40 at the low end, to a maximum of €318,000. The second and third highest values are €256,000 and €105,000 respectively. These top three figures were obtained by firms which had raised €420,000, €1137,460 & €142,500 in sales respectively.

Figure 9 Distribution of realised efficiencies for Data Pitch participants during the accelerator period



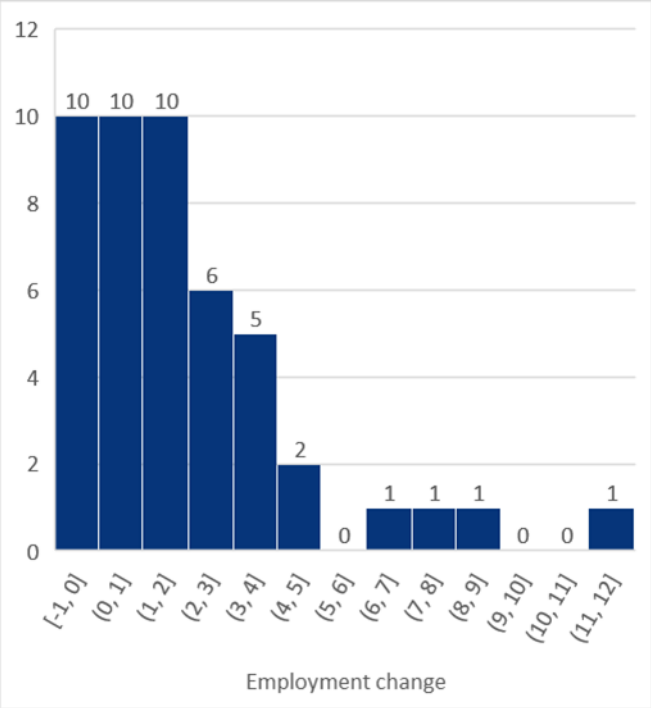
Note: Y-axis refers to number of start-ups, total N = 47.

Source: Bi-weekly monitoring updates

38 (81%) start-ups experienced a net positive change in employment. 10 firms increased their net employment by 1, another 10 firms by 2 and 17 by 3 or more, over the accelerator period. Only one firm reduced their employee headcount (-1). The highest changes in employment amounted to 12, 9 and 8, respectively.

Of these three firms, two had greatly exceeded their first- and second-year employment forecast figures, with the highest increase of 12 raising their total employment to 14 (three away from their one-year forecast of 17).

Figure 10 Distribution of employment changes for Data Pitch participants during the accelerator period



Note: Y-axis refers to number of start-ups, total N = 47.

Source: Bi-weekly monitoring updates

Twenty-five of the 47 start-ups (53%) had secured funding prior to entering Data Pitch. This funding came from sources such as other accelerator programmes, angel investors and venture capital. The three highest levels of prior funding amount to $2.3 million, $2 million and $2 million. For 22 start-ups, Data Pitch was their first source of funding.

3.5.1 Cohort 1 – Post acceleration performance

In the six months since completing the programme, start-ups from Cohort 1 have continued to increase their revenue, investment and employment figures, as shown in the table below.

Table 9 Cohort 1 post-acceleration performance (6 months after the programme)

Indicator	Observations	Mean	Minimum	Maximum	Total
Sales	17	134,379	0	1,200,000	2,284,438
Investments	17	71,334	0	665,000	1,212,679
Employment	17	1	-3	6	21

Note: Values are in euros (other than Employment). Investments include Grants, BAs, VCs etc.

Source: 6 months progress update

On average, each start-up attracted sales of just under €135,000 and investments to the value of just over €71,000. Total employment has increased by 21, with only three firms experiencing a reduction in employment (respectively -3, -2 & -1). 6 firms (35%) reported no sales, 10 firms (59%) reported no investment and 4 firms (24%) reported no net change in employment.

At 12 months after the end of the programme, cohort 1 was surveyed again on their progress in the six months since the 6 months progress update. 9 responses were obtained.

Table 10 Cohort 1 post acceleration performance (6 - 12 months after the programme)

Indicator	Observations	Mean	Minimum	Maximum	Total
Sales	9	91,529	0	480,000	825,000
Investments[a]	11	369,091 – 737,727	0	1,700,000 - 5,000,000	4,060,000 – 8,060,000
Employment[b]	5	1	-2	4	5

Note: Values are in euros (other than Employment). Investments include Grants, BAs, VCs etc.

[a] One firm reported investments of between 1 and 5 million euros; a range has been provided in summary calculations to reflect this.
[b] Two employment change figures are given by firms as having 'increased'. The calculations here exclude these observations and therefore underestimate the true net employment change.

Source: 12 months progress update; investment details for 2 start-ups received from the start-ups directly

Of the firms which filled in this survey, 4 reported no sales, 6 reported no additional investments and 1 reported zero net change in employment. The average increase in investment ranges from €369,000 to €732,727 whereas the average increase in sales amounts to €91,529. Overall, employment has increased by at least 5.

Four of these start-ups continued to obtain revenue over both six-month periods; three start-ups had not raised any sales across both six-month periods. One start-up out of the nine who responded in both surveys reported a reduction in sales from €8,000 in the first period, to zero in the second. The highest increase in sales is an increase from €200,000 to €480,000.

Five of the start-ups which reported investment figures in this period received zero in additional investment over this period. Three start-ups which raised zero in sales did raise investment in either period (with one of these firms raising €1.3 million in the second period). Four firms which raised sales revenues did not receive any additional investment. This may be an indication of the effects of company maturity, as start-ups which are pre-revenue may be focused on raising investment before the release of a commercial product. Additionally, five of the start-ups which reported in both periods reported an increase in employment over both periods, with only one start-up reporting an employment loss.

3.5.2 Return on Investment and additional investments leveraged from Data Pitch

The performance of the start-ups can be compared with the financial resources provided to the Data Pitch programme. Table 111 below shows the cumulative Return on Investment (ROI) and leveraged investments generated by the programme.

ROI is defined as sales generated by start-ups as percentage of the total monetary resource made available to Data Pitch, including but not limited to the funding provided to start-ups[38]. Sales are cumulative. That is, sales defined for the ROI in the 6 months following the programme include sales generated during the programme. Similarly, sales for the ROI in the 6 to 12 months following the programme include both the preceding periods.

Leveraged investments measure the additional investments attracted by the start-ups and enabled by Data Pitch funding. This is defined as additional investment attracted since joining Data Pitch divided by the total monetary resources available to Data Pitch. As with sales, additional investments are cumulative.

[38] About €7.8 million in total.

Data on post-acceleration performance is only available for Cohort 1. The ROI and leveraged investment figures account for this by appropriately adjusting the total Data Pitch funding based on the number of start-ups for which there is data.

Table 11 Return on Investment and leveraged investment

	During the programme	6 months following the programme	12 months following the programme
Return on Investment	23%	91%	103%
Leveraged investment	50%	82%	278%

Note: data on 6 and 12 months after the programme are based on data from Cohort 1 only. The column "6 months following the programme" uses data for all of Cohort 1, whereas the column "12 months following the programme" uses data for a subset for which data is available. The figures account for this by adjusting the investment provided through Data Pitch based on the number of start-ups for which data is available.

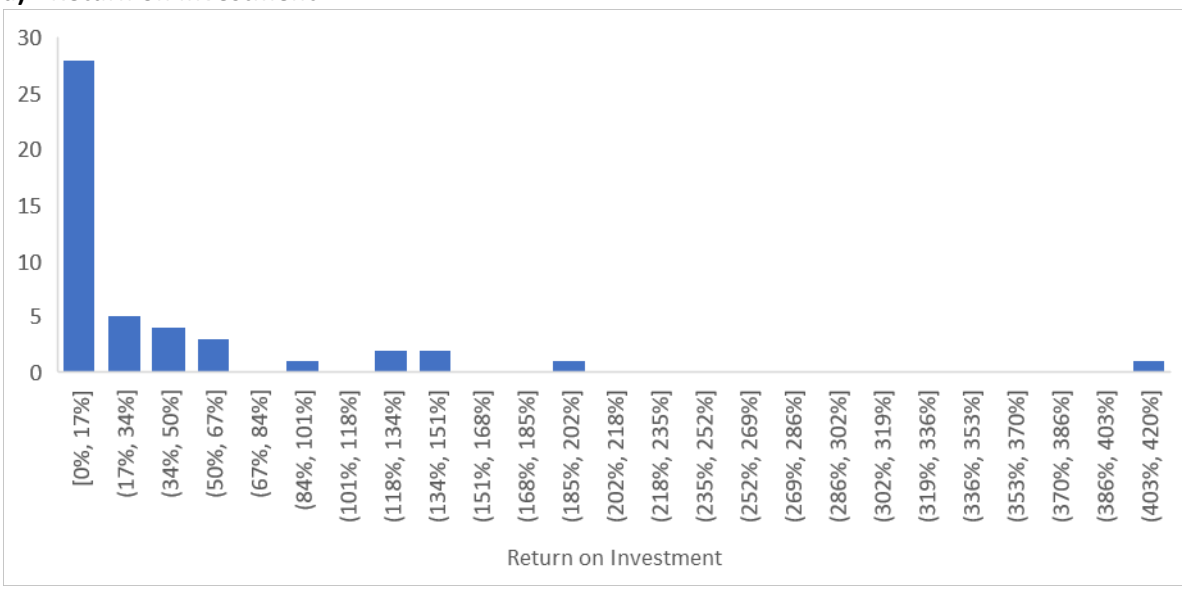
Source: Bi-weekly monitoring updates, 6 months progress update, 12 months progress update; investment details for 2 start-ups received from the start-ups directly

Both ROI and leveraged investment are already substantial during the programme. By the end of the programme, the total Data Pitch resources already attracted 50% equivalent value from other investment opportunities. One year after the end of the programme, ROI already exceeded 100% and leveraged investment was already approaching 300%.

These figures hide some variety across start-ups. The figure below shows the histogram of start-up-specific ROI and leveraged investment generated during the programme[39]. These numbers are based on sales and additional investment divided by €100,000, the maximum investment available to individual start-ups in Data Pitch.[40]

Figure 11 ROI and leveraged investment during the programme per start-up

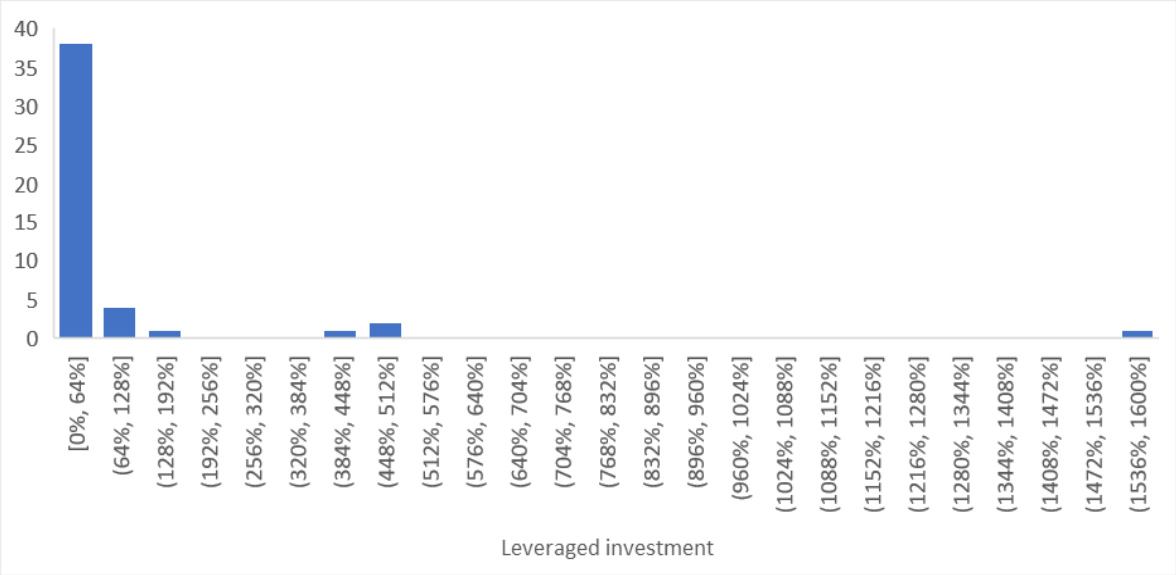
a) Return on Investment



[39] Figures for start-up-specific ROI and leveraged investment 6 and 12 months following the programme show similar qualitative patterns.

[40] The average funding received by start-ups is somewhat lower than €100,000 as not all start-ups used the maximum available funding. Therefore, these figures show conservative estimates. Furthermore, the figures do not include additional monetary resources available to Data Pitch. The use of these resources cannot be attributed to specific start-ups.

b) Leveraged investment



Note: y-axis shows number of start-up within each bin.

Source: Bi-weekly monitoring updates

For both the ROI and leveraged investment, a small number of start-ups manage to attract considerable sales or additional investment. A substantial proportion, however, has no or low ROI and leveraged investment. This follows expectations regarding start-ups and acceleration. Some accelerated businesses will be extremely profitable and show high ROI. Other start-ups may become sustainable but will not outgrow SME status.

3.6 Use of data

Data Pitch enabled start-ups to access data they would otherwise not have been able to access. Through the data provider challenges Data Pitch opened up data from data providers that would not be available for use by start-ups otherwise.

The main objectives of start-ups in Data Pitch was to make predictions with or identify patterns in data. The majority used machine learning to do this, especially start-ups in the sector challenges. Start-ups rated their solutions as moderately unique and innovative.

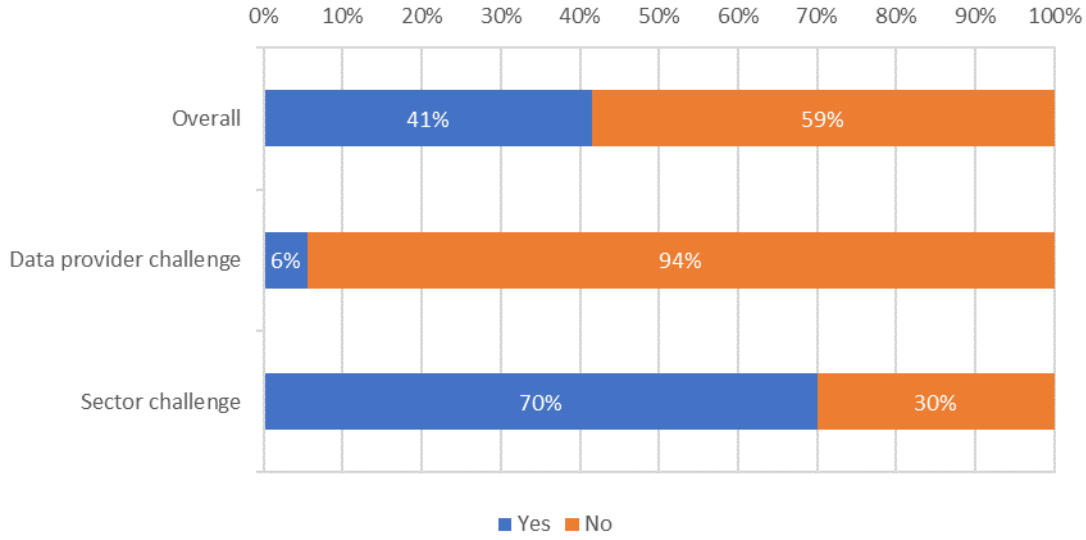
Start-ups in the data provider challenges had closer interactions with their data provider and were more likely to have pivoted their idea from their initial proposal.

3.6.1 Data shared in Data Pitch

Data Pitch enables data access

A key objective of Data Pitch is to facilitate data-driven open innovation by enabling access to data for start-ups and SMEs. Data Pitch clearly achieves this. As Figure 12 shows, **the majority of start-ups would not have been able to access the same or similar data without Data Pitch.** This is especially striking for start-ups in the data provider challenges. The figure shows that start-ups in the sector challenges would have found it easier to access data outside of the Data Pitch programme. These start-ups typically had an already established relationship with their data provider, fostering data access.

Figure 12 Would you be able to access similar data outside Data Pitch? % of respondents



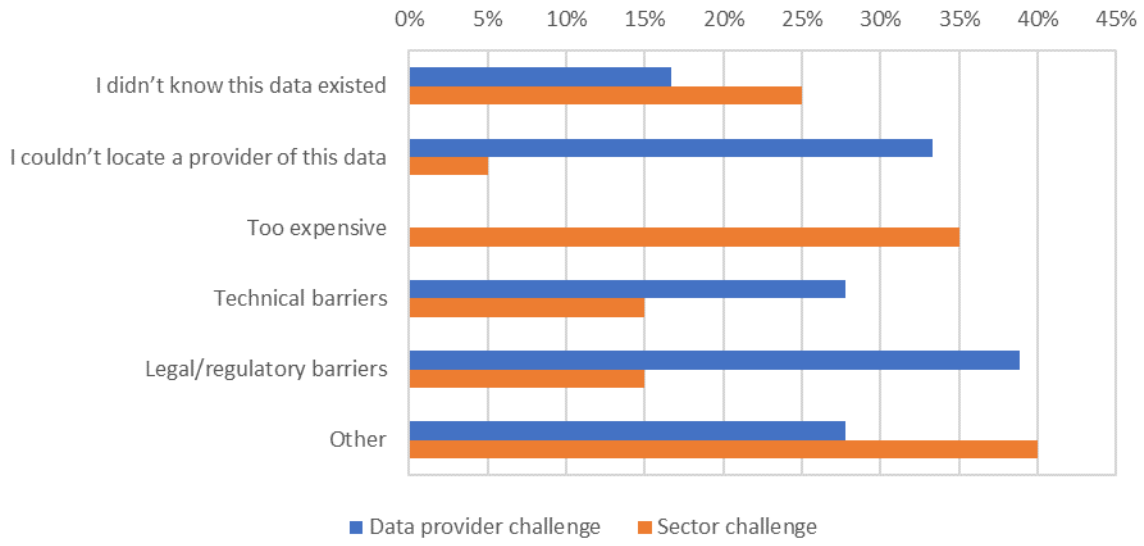
Q4: Without Data Pitch, would you have been able to access the same data (or equivalent data that would enable you to implement the same solution)?

N=41 for total; N=18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

The barriers to data access vary per challenge type. Start-ups in the data provider challenges see regulation as the greatest barrier to access. Start-ups in the data provider challenges further noted that the inability to find data providers would have stopped from accessing data. This suggests that Data Pitch managed to co-operate with the right kind of data providers, whose data would normally not be accessible. Businesses in the sector challenges, on the other hand, are more held back by the cost of data, which may reflect greater exposure to existing commercial datasets.

Figure 13 Barriers to data access; % of respondents



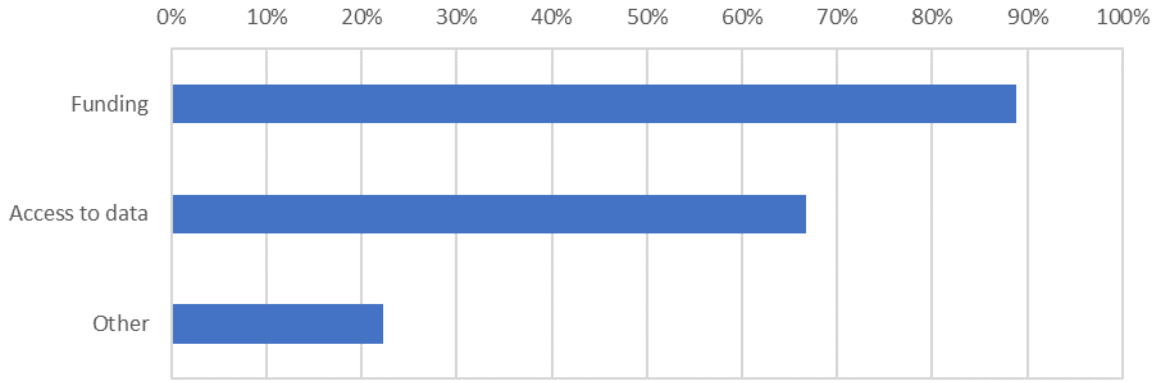
Q5: What would prevent you from accessing the same data (or equivalent data that would enable you to implement the same solution)?

N=18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

Unsuccessful applicants also confirmed that access to data was an important reason for applying to Data Pitch, albeit less important than the funding.

Figure 14 Reasons for applying to Data Pitch; % of respondents (unsuccessful applicants)



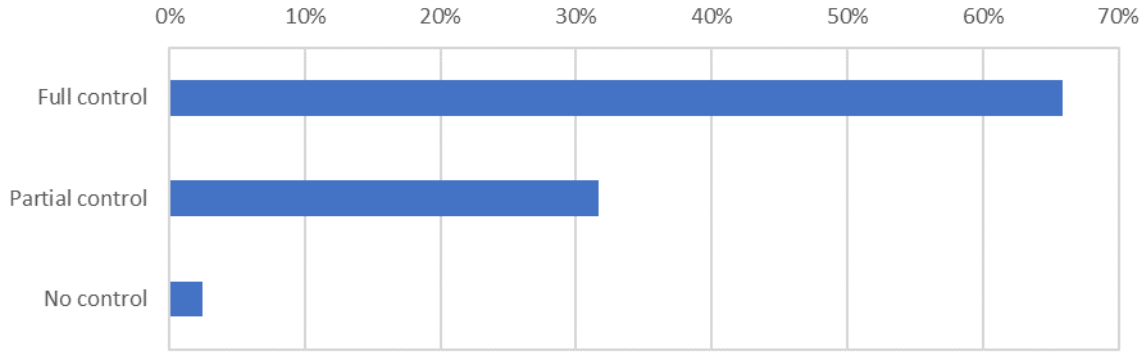
Q6: Why did you apply to Data Pitch? Select all that apply.

N = 9

Source: Unsuccessful applicants survey

Once the data was received, start-ups typically had full control[41] over the data when building the solution. Only one start-up reported that they had not had at least partial control over the data.

Figure 15 Control over the use of data; % of respondents



Q14: How much control do you have over the data when building your solution?

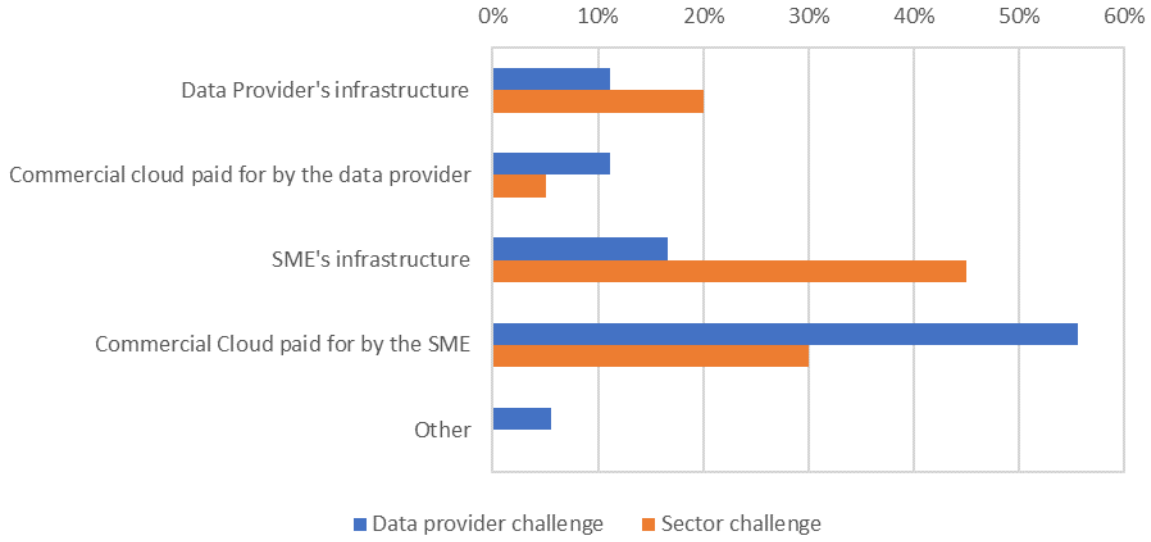
N = 41

Source: Successful applicants survey

Similarly, data was typically stored under the control of the start-up. Start-ups in the data provider challenges were more likely to use a commercial cloud service paid for by themselves. Firms in the sector challenges, on the other hand, were more likely to store data within their own infrastructure. This difference may exist because the relationship between the start-up and data provider may have been between data provider and sector challenge. In the latter, there typically was an already established relationship. This relationship may lead to more trust of the data provider in the start-up's infrastructure.

[41] The notion of full and partial control was not defined in the survey questionnaire but illustrated with an example. Full control was illustrated as "e.g. full copy of data freely available to me" and partial control was illustrated as "e.g. data called through API when required".

Figure 16 Location of data storage; % of respondents



Q13: Where is the data stored?

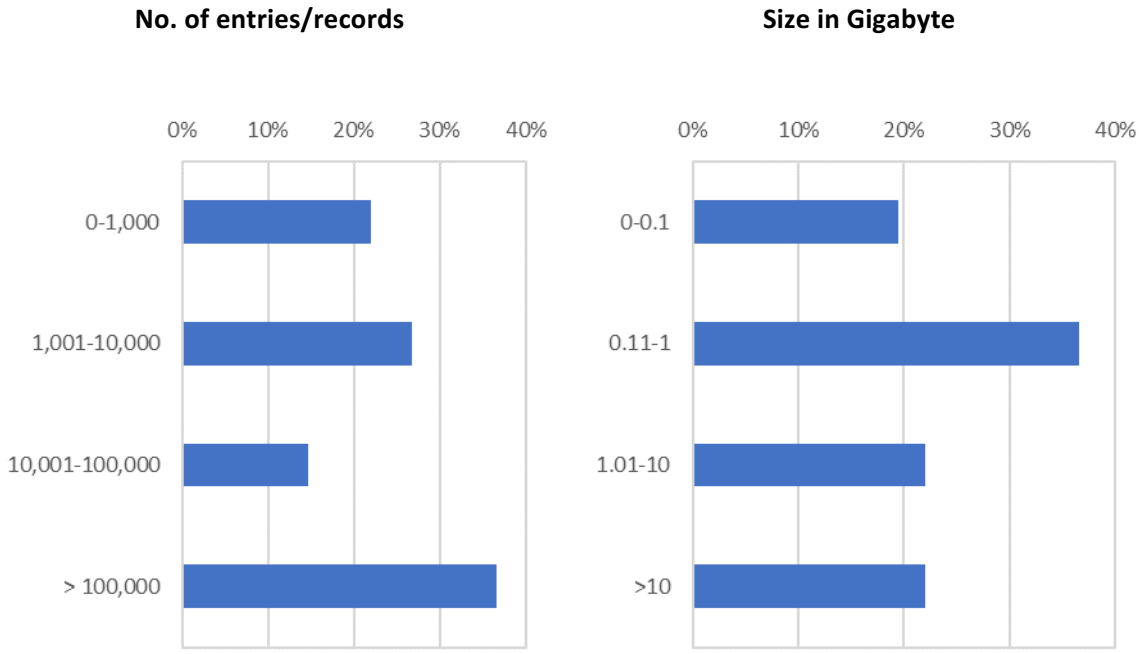
N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

Characteristics of data shared in Data Pitch

The data shared through Data Pitch does typically not constitute so-called Big Data, i.e. large volumes of data that require dedicated, high-end infrastructures to process and analyse. Over 75% of successful start-ups received 10GB of data or less during the programme. Similarly, respondents typically received fewer than 100,000 individual data items or records.

Figure 17 Size of datasets shared by the data provider



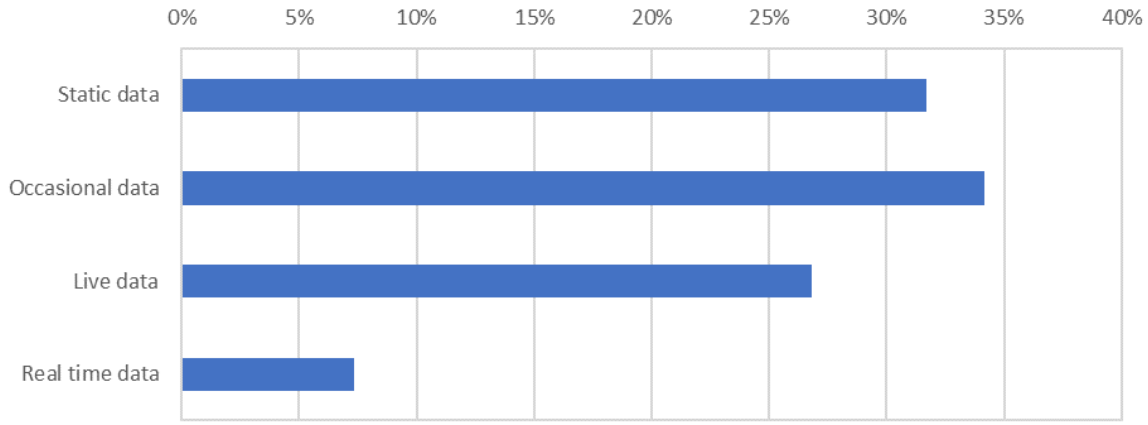
Q7: How large is the dataset that you obtained from your main (partnered) data provider for use in Data Pitch?

N = 41

Source: Successful applicants survey

Most start-ups received either static data (not being updated) or occasional data (updated infrequently). Some received live data or real-time data. The mature commercial solutions – discussed in more depth below – will move towards using live data in the future, although few will be using real time data.

Figure 18 Periodicity of data shared during the programme; % of respondents



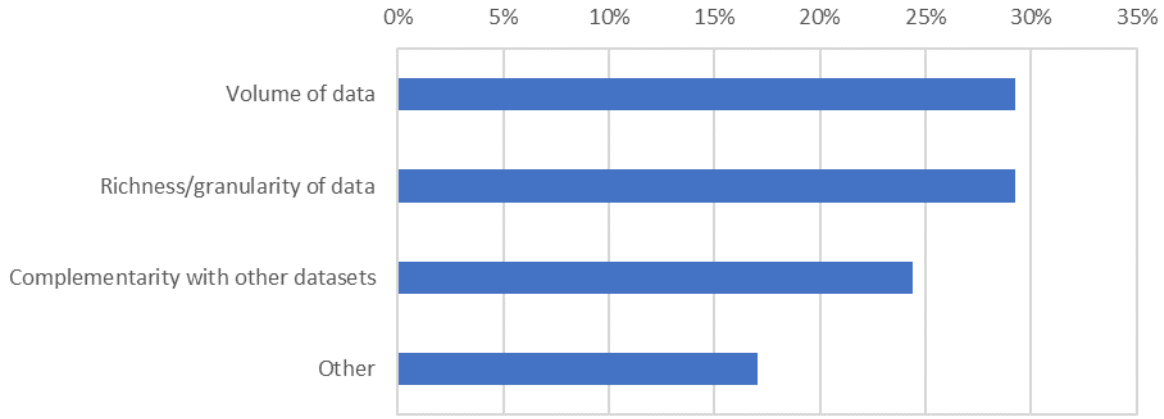
Q15: What is the update frequency (periodicity) of the data used in your solution? During the acceleration period. N = 41

Source: Successful applicants survey

Start-ups believe that both the volume and richness (or granularity) of the data was the most important for their solution. There is also some supporting evidence that the ability to combine data

with other datasets is beneficial. 63% of start-ups used 3 or more datasets in their solution (56% in the data provider challenges and 75%; see Figure 20).

Figure 19 Most important characteristics of data; % of respondents

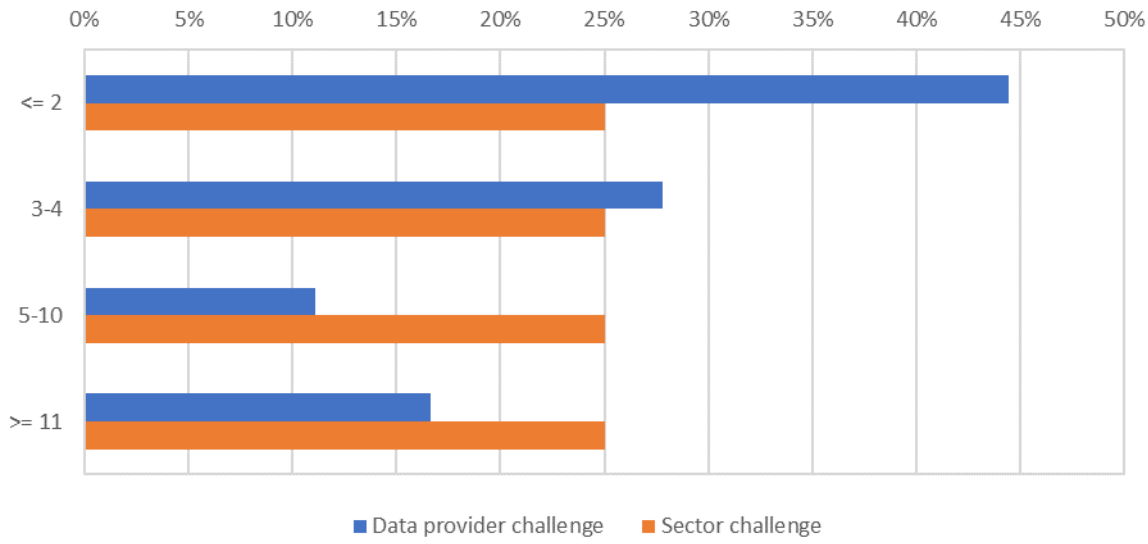


Q6: Which characteristics of the data used in Data Pitch are the most important for your solution?

N = 41

Source: Successful applicants survey

Figure 20 Number of datasets used in the solution; % of respondents



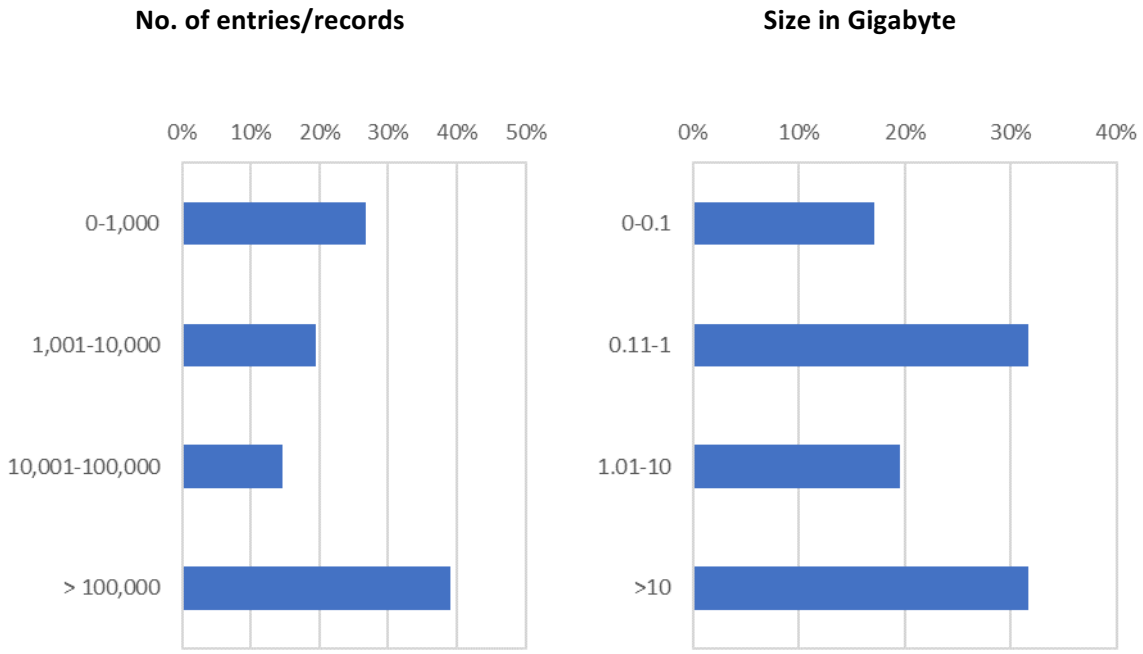
Q3: How many datasets do you use in your solution? Please provide a total number of open, closed and self-generated datasets.

N = 18 for data provider challenge; N = 20 for sector challenge.

Source: Successful applicants survey

Even though start-ups note that volume is important, they do not use Big Data in their solution. In line with the data shared through the programme, start-ups typically use less than 10GB in their solution. On average, the number of entries in the ultimate solution is larger than the data shared through the programme but the size in GB is somewhat smaller. This suggests that not all attributes of the data shared through the programme are being used.

Figure 21 Size of datasets used in the solution



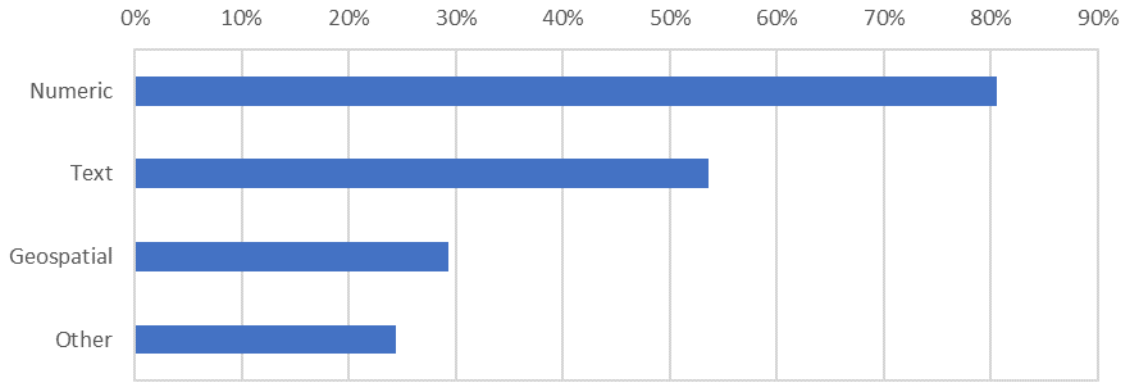
Q9: How large is the dataset that your solution uses in total? (I.e. any dataset(s) provided for Data Pitch + any dataset(s) you collected yourself or obtained from other data providers outside Data Pitch)

N = 41a

Source: Successful applicants survey

Most start-ups use numeric data. A substantial portion use text or geospatial data. Start-ups in the sector challenges were more likely to go beyond numeric and text data, and also used video and graphical data. The main types of data may explain the relatively small size of the datasets used in the ultimate solutions. Numeric and text data tends to take up less space than, for example, images and video.

Figure 22 Main type of data used in solution; % of respondents



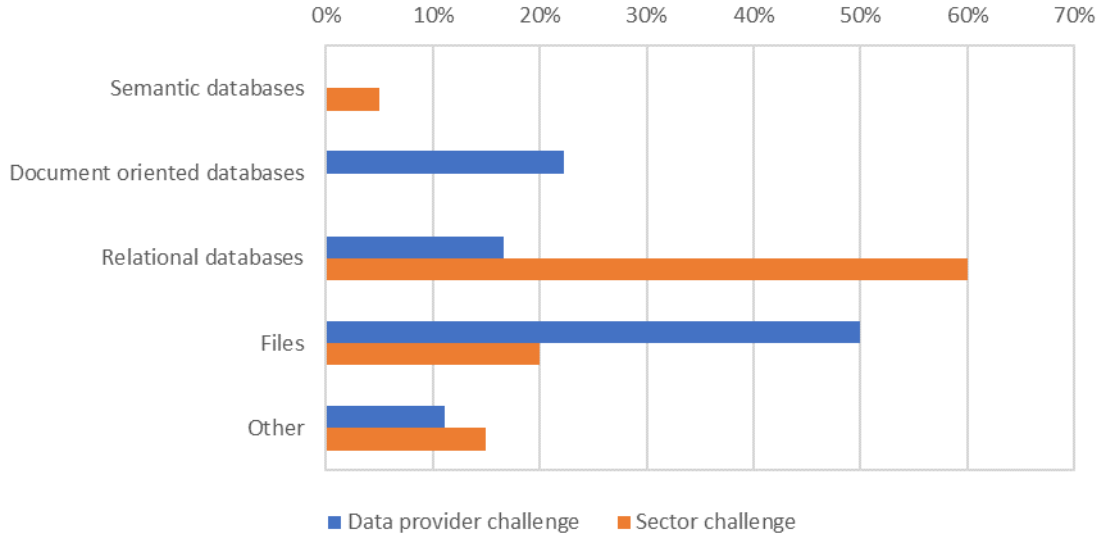
Q11: Please indicate the main type of data used in your solution, select all that apply.

N = 41

Source: Successful applicants survey

Start-ups in the data provider challenges tend to store data in files, whereas start-ups in the sector challenges tend to use relational databases. This may reflect a more ad-hoc approach to data sharing by the data providers in the provider challenges.

Figure 23 Data storage type; % of respondents



Q12: How is this data stored?

N = 18 for data provider challenge; N = 20 for sector challenge

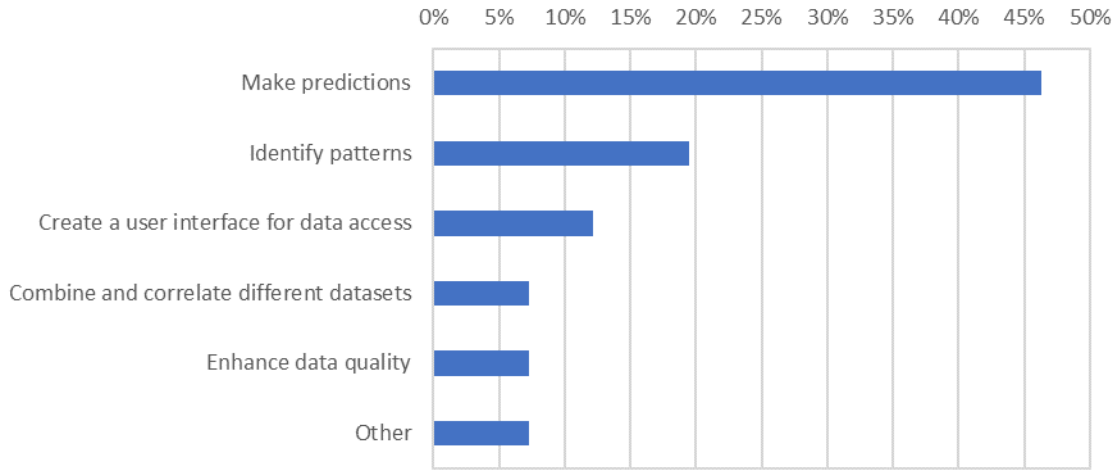
Source: Successful applicants survey

Overall, the data shared in Data Pitch appears to be at the lower end of the spectrum in terms of complexity (data types, volume and update frequency). While this says nothing about the complexity of the data itself in terms of the difficulty of making it useful in a business process, it shows that the required data engineering and data science skills did not need to include knowledge about dealing with "big data".

3.6.2 Data Pitch solutions

The plurality of solutions generated through Data Pitch use data to make predictions, followed by seeking to identify patterns or create a user interface. The first two are clear Machine Learning applications.

Figure 24 Main objective of the solution; % of respondents



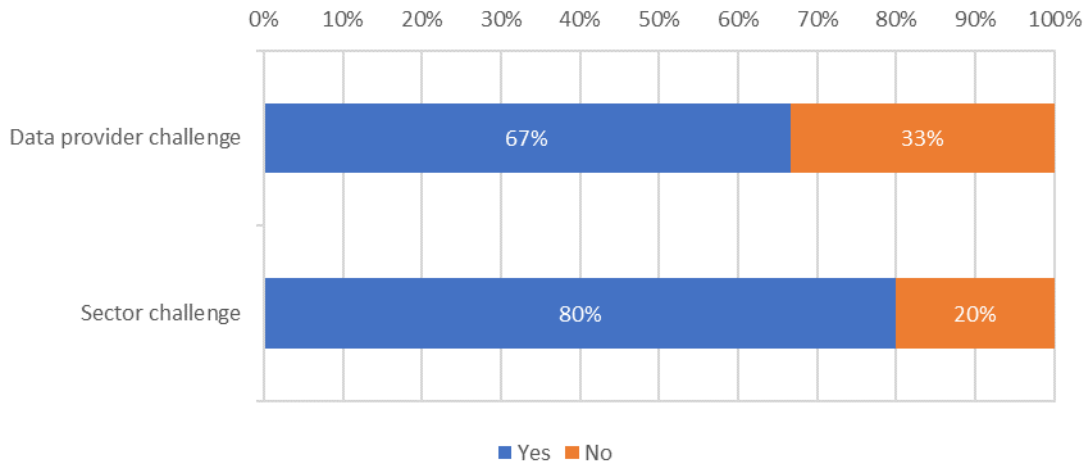
Q19: What is the primary technical objective of your solution?

N = 41

Source: Successful applicants survey

Consequently, the majority, but not all, start-ups report using Machine Learning in their solutions. There is a clear distinction between the challenge types. Start-ups in the data provider challenges tend to use Machine Learning less and tend to use regression and clustering algorithms. Start-ups in the sector challenges are more prone to use deep learning algorithms. Again, this may reflect the fact that participants in the provider challenges had less time to prepare their approaches, whereas start-ups in the sector challenges may have worked with the relevant data and methods for longer periods of time. It may also be the case that data providers in the data provider challenges were more cautious and sought more mainstream solutions, as many of these data providers are still in the early phase of experimenting with open innovation. These data providers may also have been more cautious as they had to work with start-ups unfamiliar to them.

Figure 25 Use of Machine Learning; % of respondents



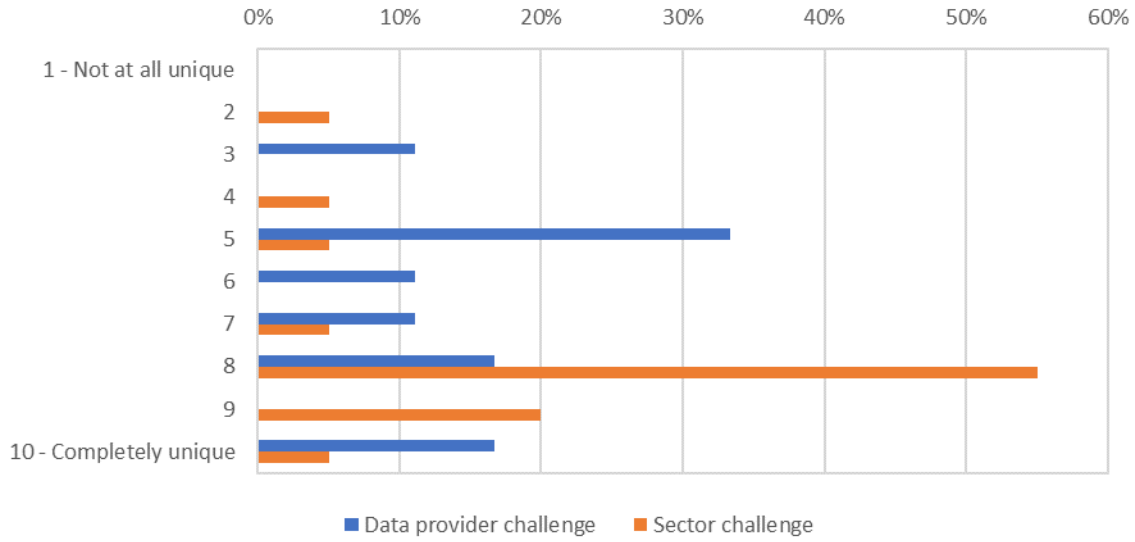
Q20: Does your solution use Machine Learning?

N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

In relation to the solution, the survey asked about the perceived uniqueness and innovativeness of the solution. Start-ups believe their solution to be moderately unique. Businesses in the sector challenges rate their solution as more unique. This could be a result of participants in the provider challenges working on a narrower, more defined, set of business problems.

Figure 26 How unique is your solution? % of respondents



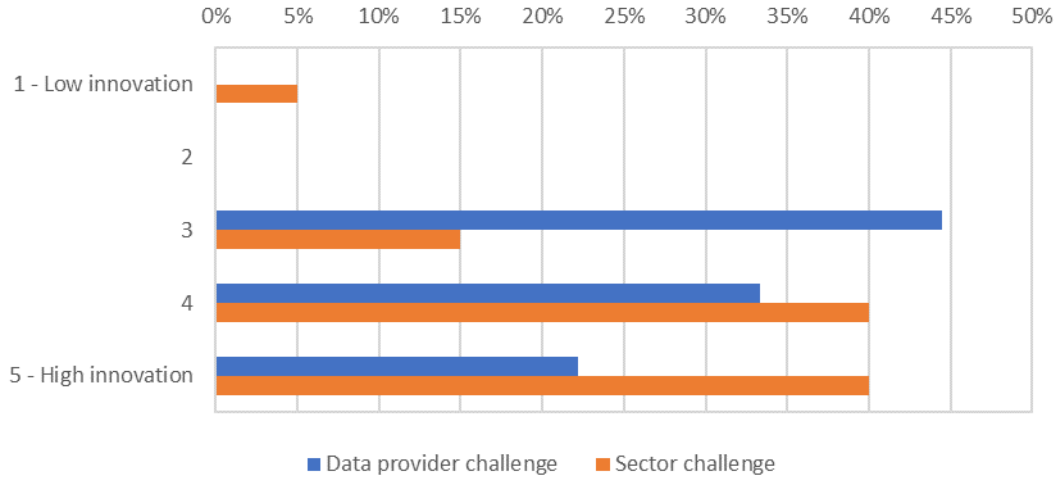
Q30: On a scale to 1-10, how unique is the product that your solution provides to your customers? Are there similar types of products out there, or is this a one-of-a-kind? (An answer of 1 signifies the product is 'not at all unique' and an answer of 10 signifies the product is 'completely unique').

N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

These findings are mirrored in perceived innovativeness of the solution. Here, start-ups in the sector challenges tend to rate their solution more innovative. Not surprisingly, the perceived uniqueness and innovativeness of solutions are highly correlated.

Figure 27 How innovative is your solution? % of respondents



Q31: On a scale of 1-5, how innovative is your solution? (An answer of 1 signifies low innovation, an answer of 5 signifies high innovation).

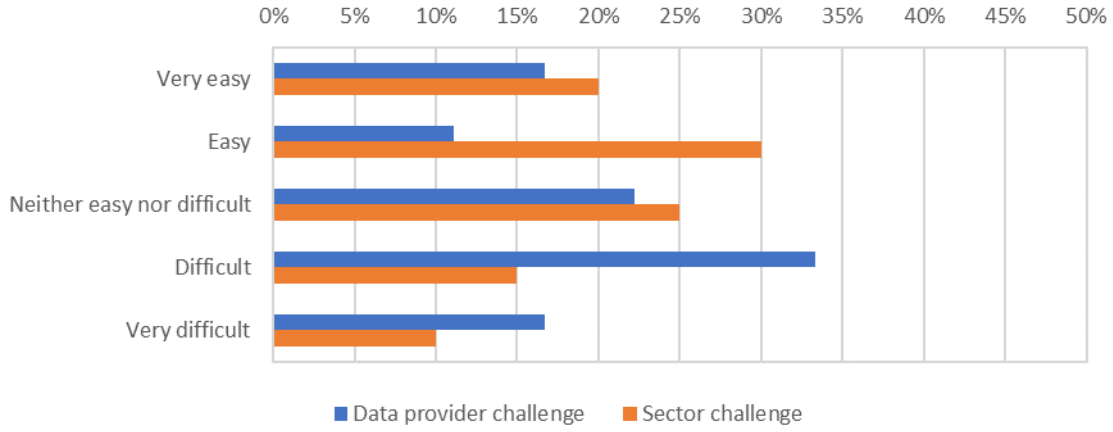
N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

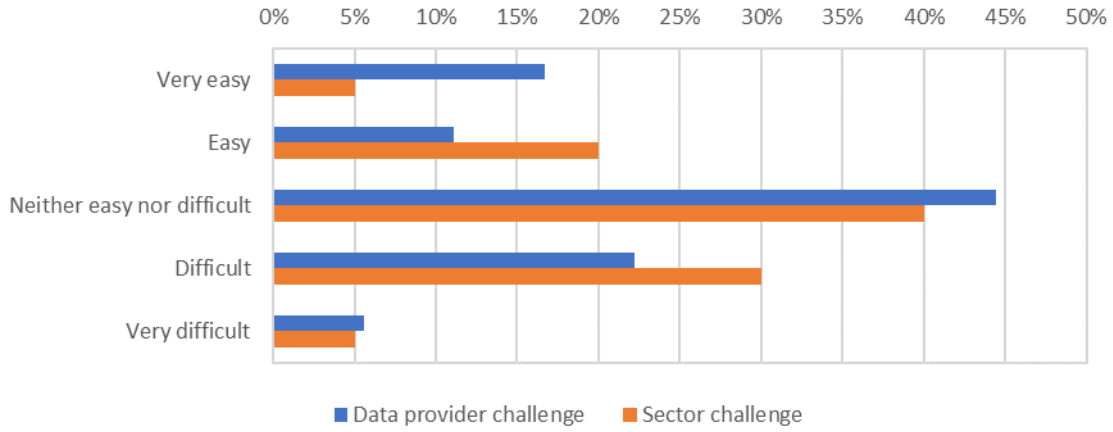
Figure 28 shows the perceived difficulty of start-ups to access data, get it into the right shape (data engineering) and building the solution.[42] Overall, data access was considered the easiest and building the solution was considered the most difficult.

Figure 28 Perceived difficulty of access, engineering and building; % of respondents

a) Data access

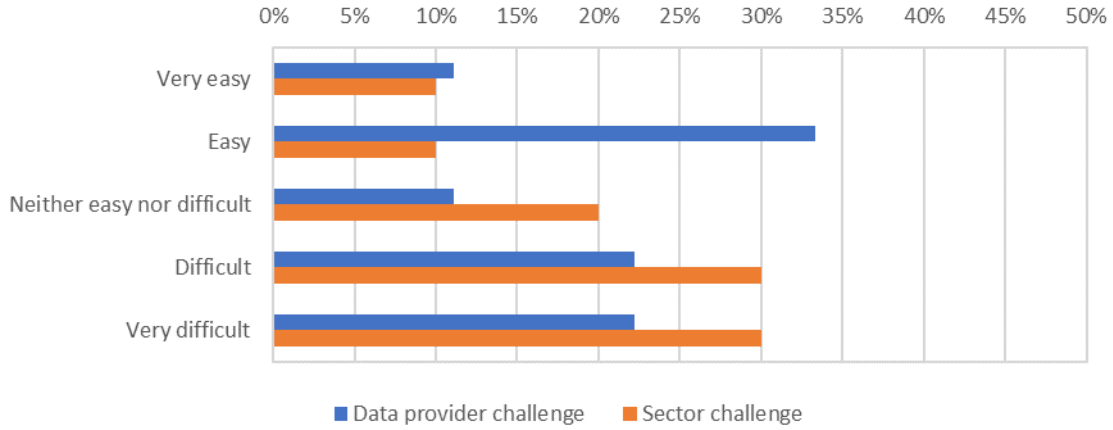


b) Data engineering



[42] Note that these figures refer to access to and engineering of data shared with start-ups through the Data Pitch programme. These figures do not refer to a general ability to access or engineer data.

c) Building the solution



Q18: For each of the below, please rate the difficulty you had during the acceleration period: (1 signifying easiest, 5 signifying hardest)
N = 18 for data provider challenge; N = 20 for sector challenge
Source: Successful applicants survey

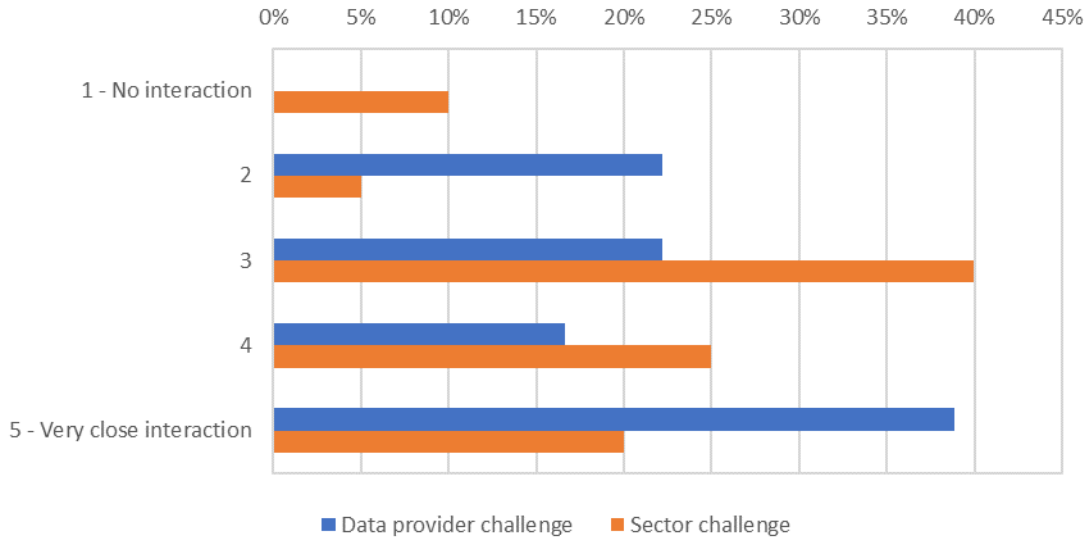
Start-ups in the data provider challenges found accessing data more difficult than start-ups in sector challenges, presumably because start-ups in the sector challenges had a pre-established relationship with their data provider. On the other hand, start-ups in the data provider challenges had an easier time engineering the data and building the solution. In part this likely reflects the methods adopted (see Figure 25).

Part of the perceived difficulties can be attributed to the interaction between data provider and start-up.[43]

As shown in Figure 29, most start-ups report that they had at least some interaction with their data provider. Start-ups in the data provider challenges were more closely interlinked with their data provider than the start-ups in the sector challenges. The results also show that start-ups that interacted more closely with the data provider matched through Data Pitch interacted less with data providers outside of the programme.

[43] The interaction between SME and data provider is discussed in more detail below.

Figure 29 Interaction between start-up and data provider; % of respondents



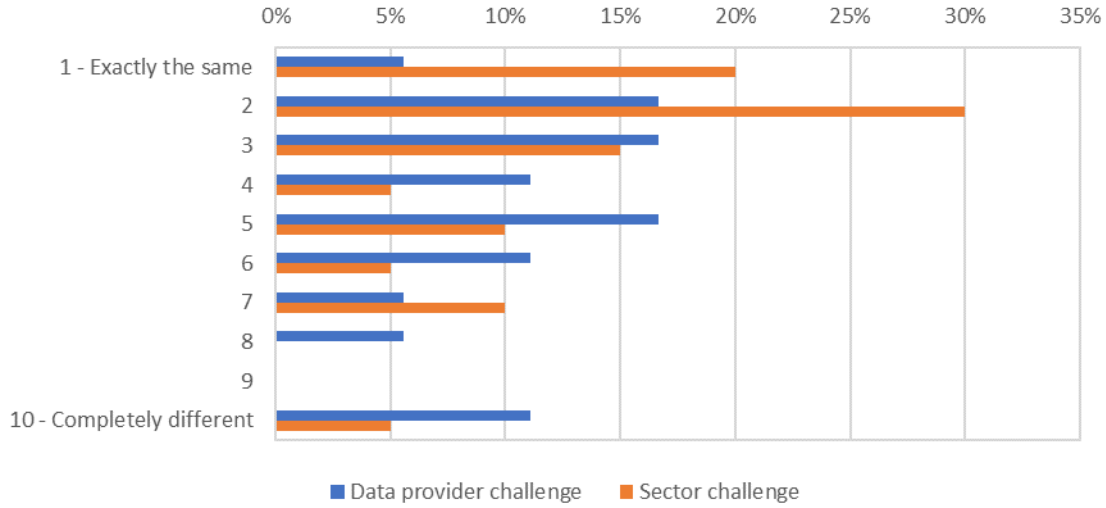
Q17: How closely did you interact with your partnered Data Provider? (1 signifies 'No interaction' and 5 signifies 'Very close interaction')
N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

Start-ups that had closer interactions with their data provider found accessing data easier. This correlation is especially strong among the start-ups in the data provider challenges. On the other hand, building the solution was perceived as harder by start-ups with closer interactions with their provider. Possibly, a close interaction between start-up and data provider allowed the start-up to better understand the provider's businesses need, which meant that preconceived ideas about the solution had to be abandoned.

Indeed, most start-ups had to change their initial proposal at least somewhat during the programme (see Figure 30). Most changes occurred due to business reasons, highlighting that a better understanding of the business needs may indeed have driven changes in the solution. This notion is further strengthened by the fact that start-ups that had a closer interaction with their data provider also reported that their ultimate solution is more different from their initial proposal.

Figure 30 How different is the solution from the initial proposal? % of respondents



Q23: How different is your solution now from the idea you had at the start of your involvement in Data Pitch? (An answer of 1 signifies your solution matches the initial proposal exactly; an answer of 10 signifies that the solution is completely different from the proposal).

N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

3.7 Drivers of outcomes

Start-ups generated €599,432 in sales and attracted €338,862 in additional investment per Gigabyte shared with them through Data Pitch during the programme. There is, however, no linear relationship between the size of datasets and the value generated.

There is considerable variation between outcomes for start-ups operating in the same sector. Tentatively, it seems that health start-ups still rely on investment, start-ups working in heavy industry are maturing to be sales-driven and start-ups in the financial sector show most employment growth.

The use of machine learning in the solution impacts the ability to attract investment. Perceived ability to scale a solution is also related to the ability (or willingness) to attract funding, with more scalable solutions attracting more investment.

The programme was delivered efficiently, fostered cooperation between start-ups and data providers and enabled market entry by matching start-ups with potential customers. Data Pitch worked particularly well for start-ups with pre-existing ideas or working products.

Data Pitch tracked outcomes of start-ups during the programme. This section highlights some potential drivers of these outcomes.[44]

[44] Note that the data in this section is based on the 41 start-ups that completed the successful applicants survey, rather all 47 funded firms.

3.7.1 Shared data

To understand potential drivers of outcomes, it helps to split start-ups into groups based on certain characteristics. An obvious first split is the ability to access data outside of Data Pitch.

As noted above, a main aim of Data Pitch was to enable data sharing and the survey results showed that a large portion of start-ups would not have been able to access similar data without Data Pitch. Access to data does seem to influence the ability of start-ups to attract additional funding. Firms that reported they could access data attracted, on average, €177,059 in additional funding whereas firms that are not able to access data attracted €36,044 on average. The ability to access data may be a sign of maturity, explaining why this difference exists.

Moreover, the size of datasets matters. On average, start-ups generate €599,432 in sales and €338,862 in investment **per GB of data shared** with them through Data Pitch. This partly stems from the fact that the majority of start-ups used a relatively small amount of data in their solution. But it shows that much value can be generated from relatively little.

However, there does not seem to be a constant or even uniformly positive relationship between value generated from a dataset and its size. Within Data Pitch, it seems that the start-ups using "middle-sized" datasets (0.11 to 1GB) managed to attract most sales and additional investment, whereas start-ups using smaller or larger datasets generated less value. However, the relationship between dataset value and size is complex and will depend on specific circumstances.

3.7.2 Customers & data access

Data for all 47 participants in Data Pitch on their intended customer base, their ability to access data and outcomes generated during and after the programme[45] were analysed to understand the interaction between customers and data access. Individual level data cannot be published in this report due to the sensitive nature of the data, but the results are discussed below.

As discussed in section 3.2, a substantial share of start-ups is active in the health, transport/mobility and financial sectors. Other sectors covered include retail, industry with large assets such as machinery (including utilities) and the food and hospitality industries.

Among the three larger sectors, data seems most accessible for the mobility sector. Start-ups working in the financial and medical sectors were more likely to report that – outside of Data Pitch – they would not have been able to access the same data. This likely reflects the sensitivity of financial and healthcare data (which is typically personal data), including the fact that data access in those sectors is governed by sector-specific regulatory frameworks in addition to general regulation relating to personal data (GDPR).

There is considerable variability in sales, investment and employment changes even within the different sectors covered by the start-ups in the programme. This makes comparison across sectors challenging. Nonetheless, a number of tentative qualitative conclusions can be drawn.

Start-ups active in the health sector typically still rely on investment, raising more in investment during and after Data Pitch than they receive in sales. In contrast, start-ups whose intended

[45] Outcome data was derived from the bi-weekly monitoring updates, the 6 months progress update and the 12 months progress update, where available.

customers operate in heavy industry and utilities are maturing away from investment and instead show good sales figures.

Start-ups active in transport/mobility tend to attract both sales and investment, although this does not translate directly into employment growth. Start-ups in the finance sector, in contrast, show more promising employment growth. Furthermore, there is a mix of start-ups relying in investment and start-ups maturing towards sales.

Start-ups with retail and food as their target markets appear weakest in terms of investment and sales performance. This may be because these are already very competitive markets, and the margins are low. Therefore, a company needs to be exceptional to attract funding and sales.

3.7.3 Solution characteristics

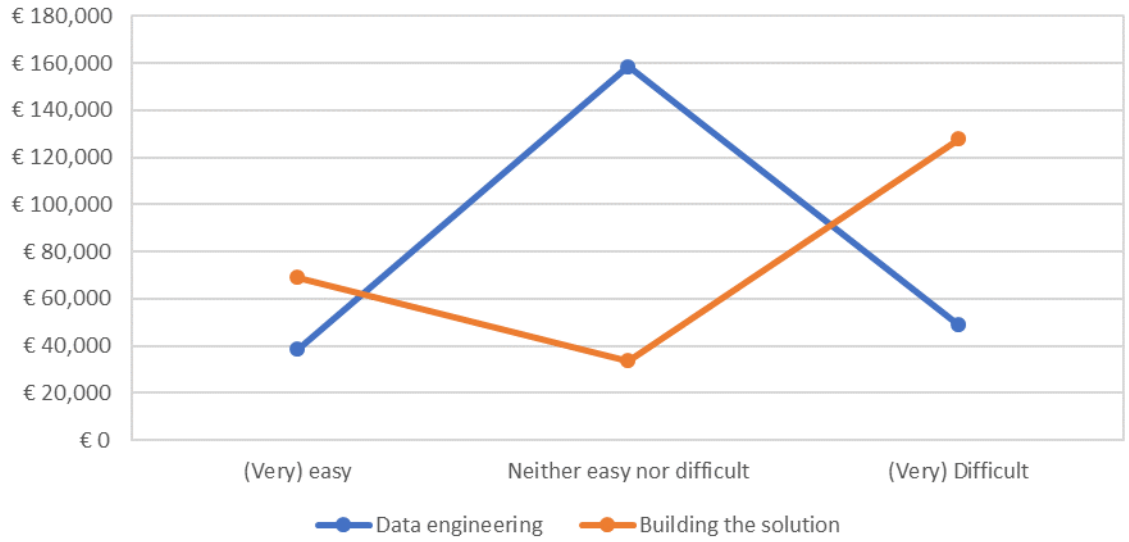
The majority of start-ups used machine learning in their solution. The outcomes tracked during the programme provide some evidence that using machine learning helps attract investment.

There is an impressive difference between start-ups that used machine learning and those that did not. Start-ups that did use machine learning attracted, on average, €126,209 in investment. Start-ups that did not receive investment, only attracted an average of €17,917. This hints towards a potential "tech bias" among investors.

The ease with which data can be engineered into shape and the ease with which a solution can be built may also impact the value generation process.[46] As Figure 31 shows, the impacts are very different for both. One may expect that the ability to attract investment improves engineering or building is (perceived to be) easier. On the other hand, investors may appreciate start-ups that realistically perceive the ease of working with data to be at least somewhat difficult. Given that there is no uniformly positive or negative relationship, both may be at play.

[46] Note that this section excludes the ease with which data can be accessed. The successful applicants survey asked specifically about the perceived ease of access, engineering and building the solution within the context of the data used in the Data Pitch programme. Perceived ease of engineering and building solutions are likely to have generalisable elements. Access on the other hand crucially depended on co-operation with the data provider. This may not generalise to other situations with other data providers.

Figure 31 Average investment per assessment of ease of data engineering and building; €

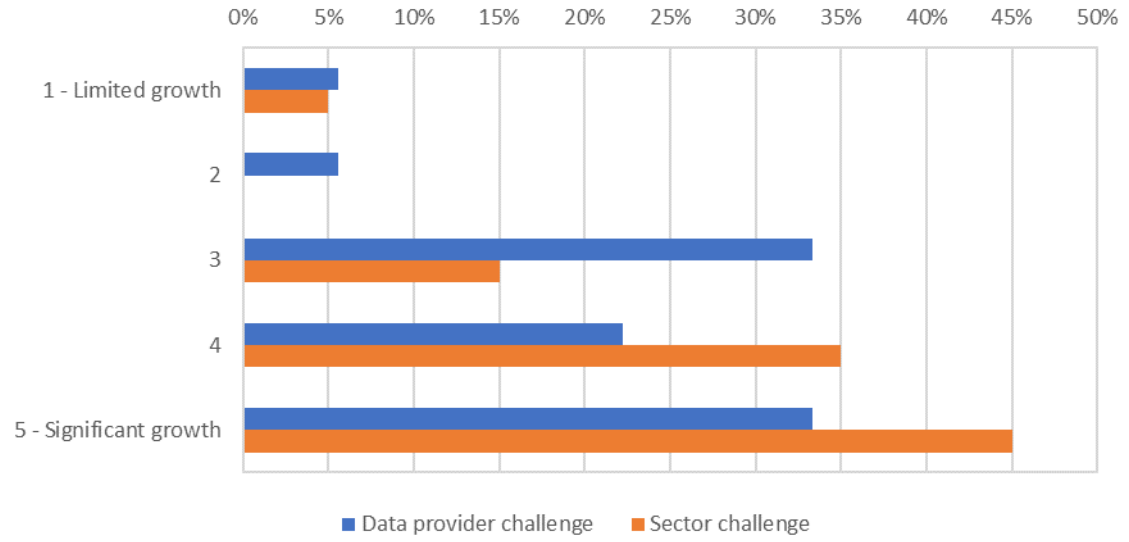


Source: Successful applicants survey and bi-weekly monitoring updates

3.7.4 Scaling potential

Most respondents believed that their core solution[47] can be scaled. Especially the start-ups in the sector challenges believed in significant growth for their product. Start-ups were even more positive about breaking into new markets, new customer groups or new application areas. More start-ups believed in significant growth opportunities for such new ground than for the core solution.

Figure 32 Scalability of start-ups' core solution; % of respondents



Q28: How do you see the scalability of your solution over the next three years? (An answer of 1 signifies 'limited growth', an answer of 5 signifies 'significant growth'). Core solution only: markets & customers as currently identified

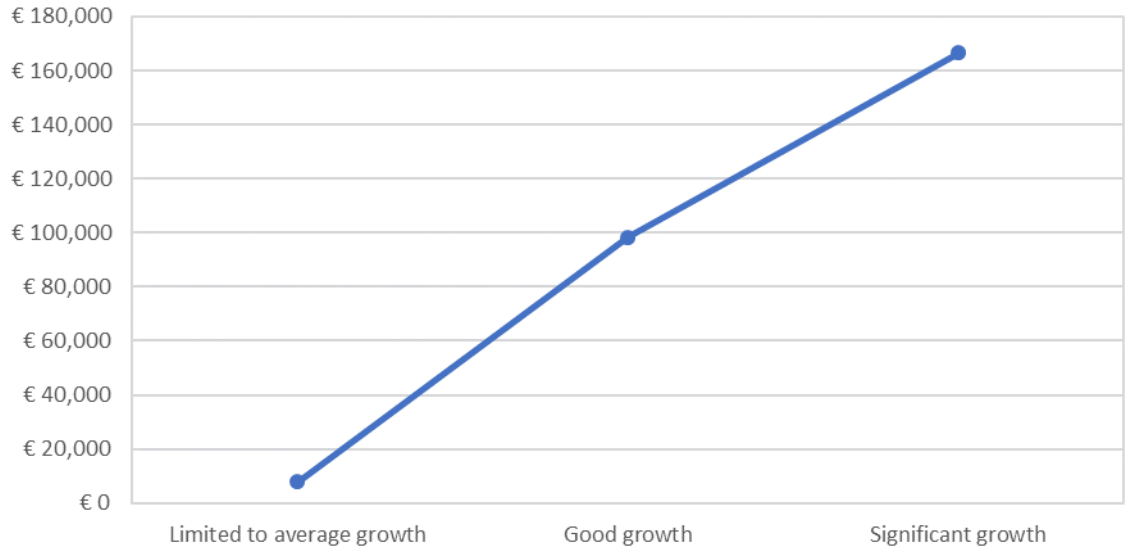
N = 18 for data provider challenge; N = 20 for sector challenge

Source: Successful applicants survey

[47] The solution developed through Data Pitch provided in an already defined market to an already defined customer base.

The (perceived) ability to scale a solution has a distinct impact on the capacity or willingness to attract funding. As shown in Figure 33, the group that perceived only limited growth or even neutral growth only attracted, on average, €7,692 in investment. The group that rated the opportunities to scale as reasonably high, attracted €98,231 on average. The group that perceive high growth in scaling their core solution attracted €166,537 on average in additional funding. In all, perceived ability to scale and attracting funding are highly correlated.

Figure 33 Average investment by assessment of opportunities to scale, €



Note: data on perceived growth is drawn from Q28.1 in the successful applicants' survey. The groupings are defined as follows: Limited to average growth: options 1, 2 and 3; Good growth; option 4; Significant growth: option 5.

Source: Successful applicants survey and bi-weekly monitoring updates

3.7.5 Participants' observations on drivers of outcomes

Several factors were identified by start-ups and data providers as making a significant contribution to the success of Data Pitch.

Outreach and recruitment

For optimal impact, it is important that a programme like Data Pitch reaches the right participants; organisations that have the motivation, skills and resources to participate in open innovation and realise solutions within the acceleration period.

Outreach activities undertaken by Data Pitch were successful in that they activated a large network of multipliers and reached relevant start-ups and data providers across Europe. Interviews suggest that previous engagement with the consortium partners and their staff was important in many cases, but the geographic spread shows no sign of bias. The interviews also confirmed that some start-ups came across Data Pitch through simple web search, which indicates that the programme was accessible to all interested start-ups.

The situation is less clear in relation to data providers. Some data providers felt being matched with several start-ups diluted the impact of the programme and put some strain on the resources they could dedicate to the programme. For future programmes, this means that there is value in making

data providers understand the resource requirements for a project like Data Pitch. This may include briefing data providers on expected capacity and ensuring that providers can provide this capacity for the duration of the programme before committing to the programme. This information exchange could be part of a pre-acceleration phase ("phase 0"). Furthermore, it may be helpful to recruit more data providers.

Project set-up

Data Pitch succeeded in providing support to a very diverse range of organisations; from small idea-stage start-ups to successful high-growth SMEs and, on the data providers' side, from large corporates to research consortia. The Data Pitch package seems to have worked particularly well for start-ups with an existing product or service that they could develop further during the acceleration period, particularly where it allowed them to access a new market or new set of clients. There is some evidence that early-stage start-ups struggled more in the data provider challenges. In part this reflects frictions in the interaction with the data providers. Early stage start-ups are likely to have less experience working with large organisations and corporations. This lack of experience may lead to a mismatch of expectations between the early-stage start-ups and data providers. Early-stage start-ups may, for instance, not be aware of the time and effort required to get approval for data-sharing within large corporations. They may, therefore, expect to receive data more quickly than a data provider can deliver.

It is also evident that participants in the provider challenges had in many ways a very different experience than those in the sector challenge, with closer interaction with the data provider and a clearer focus on open innovation.

Some data providers desired a greater focus on the data providers with a view to maximising the mutual benefits of the acceleration period. Most successful examples of co-operations between data provider and start-up showed a great deal of intrinsic motivation and prior experience in the open innovation field on the side of the data provider. In particular, data providers would have liked more interactions with other data providers. One provider noted, for instance, that a networking event involving the participating data providers could have been useful to share experiences and best practices regarding open innovation. It appears that data providers often are not mature enough to make datasets available in a way that allows immediate work on solutions to their business problems. The frictions are both technical and related to the business cases.

Several data providers remarked that Data Pitch did not particularly serve a role as a full intermediary between data providers and start-ups in terms of ironing out technical and business challenges. In the only case where real frictions arose between a start-up and a matched data provider, it was remarked that: "There should be some thinking into what to do when the data provider and SMEs incentives and goals are not strictly aligned". In this case, frictions arose from both the start-up and the data provider. On the side of the start-up, the solution was not a good fit, and the necessary pivoting was resisted by the start-up. Access to data and data quality were problematic. On the side of the data provider, the team dedicated to Data Pitch changed and were not aligned with the originally envisioned plan. However, the start-up was very young, started just before Data Pitch, with perhaps unrealistic expectations and little business experience. Furthermore, the data provider also was not experienced with open innovation.

While the programme duration is in line with other accelerators, several data providers indicated that the time was insufficient for them to realise value from the solutions. They felt that 6 months was insufficient to create a minimal viable product from scratch. Some data providers mentioned a mismatch between what start-ups think they can do in 6 months and what they can actually do. A

couple of start-ups remarked that more time would be needed to build more technically ambitious solutions and that six months was not long enough to build a viable team, a model and platform for the solution.

A couple of start-ups mentioned that the terms of the grant payments were not ideal for them, either because they would have preferred more regular payments or because they would have liked more time to use the budget (beyond the acceleration phase).

Efficient delivery

Participants were very understanding of the need for tight monitoring on the part of Data Pitch, in particular the responsibilities that come with using public funding. Only two start-ups voiced unprompted criticism of the administrative burden ("Some weeks we didn't' have anything to report"), while many more were complimentary about the efficiency of their interactions with Data Pitch ("the programme was handled perfectly, don't see any point which could be improved"). Indeed, start-ups considered the monitoring (including advisory and coaching) arrangements to be useful ("Very good idea to report bi weekly – it was very productive to help us reflect on where we stood"; "biweekly check-ins didn't really exceed 10 minutes"; "the biweekly meetings in the beginning were amazing"). Data Pitch was compared very favourably with other accelerators, especially publicly funded ones. The use of tools like FP6 is seen as a plus and shows a good understanding of the operating environment. One of the start-ups commented that Data Pitch was "more professionally run" than the another public incubator, in which they also had taken part, while another commented that both were similar in terms of administrative burden.

Collaborative open innovation

Bringing together data providers interested in open innovation and start-ups with the skills to implement innovative data-enabled solutions is in many ways the core achievement of Data Pitch. This was universally recognised by participants in the provider challenges (both data providers and start-ups). There is clear evidence that Data Pitch enabled start-ups to access unique and valuable datasets to which they would not have had access otherwise. This is corroborated by responses to the 'successful applicants' survey that clearly shows that especially start-ups in the data provider challenge would not have been able to access data similar to the data shared through Data Pitch. While some start-ups remarked that they would have liked more data, all agreed that the combination of relevant data and the real-life business needs of the data provider helped them to achieve their acceleration goals[48]. By subsidising data access, Data Pitch fills a gap in the innovation support landscape that is not addressed by other measures. For example, the cost of data can't be claimed under R&D tax credits (Coadec & The Entrepreneurs Network, 2019). Furthermore, in many cases, commercial data is not available for sale, even if funds would be available. Data sharing solutions like Data Pitch can also open up these datasets.

Data providers reported clear and substantial benefits in terms of organisational commitment to open innovation. One highlighted that data sharing has always worked well within their own field, but cross-fertilisation with other areas is the challenge they wanted to solve. Most of the data providers showed a clear understanding of the concept and its benefits ("open innovation means

[48] Bone et al. (2019) find that "start-ups which receive access to partners and customers, help in refining and testing their business model, and/or help with team formation are most likely to think the programme they participated in positively impacted their chance of success". While they caution that the types of support perceived to be important by start-ups are often not the same as those that are associated with start-ups success, they do not address the unique data-sharing function of Data Pitch in any way, which makes their results difficult to apply to Data Pitch.

working together with start-ups and academics to get in new ideas from the outside, but also sharing our knowledge to develop new solutions, approaches and innovations"; "we wanted to work with start-ups to figure out the challenges to accessing and processing and working with our data, learning how to explain and share this data was a motivation for participation").

Several data providers commented that the relationships between start-ups and data providers, and the overall success of the collaboration, depends in no small measure on the individuals involved. Therefore, the programme can be susceptible to personnel changes or setbacks. Several of the data providers reported that the persons originally driving their organisation's participations had since left, or were unable to contribute during the programme. Collaboration between start-ups and data providers seemed to be difficult in those cases.

Mix of support

Participants were near-unanimous in the view that Data Pitch worked as a package; the programme provided all the "resources necessary to structure and run the project – could not have done the project without either the support, money or data. Would not have been able to do this with just the dataset (and no money)". Funding, which was used primarily to recruit technical staff, played a key enabling role for all start-ups. However, for start-ups in the sector challenges, funding seems to have been more important than for those in the provider challenges. The sector challenges thus look a lot more like a traditional accelerator, in which funding and direct support are crucial.

Other elements of support that were seen as useful include the provision of AWS credits and mentoring and expert sessions. AWS credits were seen as crucial by several start-ups, although one would have preferred that the optional AWS scheme was extended to other providers of similar services.

Overall mentorship was seen as a useful component of Data Pitch and individual advisors and mentors received praise. However, a few start-ups fed back that the depth of expertise of the mentors was insufficient, either in terms of technical knowledge, domain knowledge, or business experience. Given the breadth of sectors, start-ups and data providers targeted by Data Pitch, a wider knowledge base for mentors would have helped.

Market access

Data Pitch helped start-ups to overcome barriers to entry to new markets. ("Data Pitch acts as even more than a shortcut, a 'new door' which wouldn't have existed without the programme, a new way in.") By overcoming information asymmetries and enabling trusted relationships between data providers and start-ups, the programme provides significantly added value compared with traditional accelerators.

European dimension

The ability of Data Pitch to connect organisations across Europe was seen as an advantage, especially for businesses outside established start-up hubs. Data Pitch enabled matches between data providers and start-ups that were otherwise unlikely either because they are located in different countries or because providers and start-ups are not aware of each other's existence. Some data providers even noted that they specifically participated in Data Pitch to work with start-ups that otherwise they would have not been aware of. However, some participants remarked that data is often location specific, so that the utility of the data may be reduced.

One start-up mentioned healthcare data in his home country is especially difficult to access, so that Data Pitch provided a great opportunity to work with this type of data. This highlights an important point that data ecosystems continue to differ along many dimensions, so that enabling data sharing across borders is a benefit in itself.

On the flip side, the virtual set-up of Data Pitch led to less close contact and reduced the level of support the data provider could provide in some cases. Some participants observed that the Data Pitch's base was "very London-centric", whereas most organisations were not based in London (both data providers and start-ups). They felt that they could have got more out of the programme if there had been more interaction. This was especially relevant for data providers. Start-ups in general saw the virtual aspect as an advantage that fit in with their way of working (several had operations in several European countries, in particular through developers working remotely).

3.8 Counterfactual analysis

Successful applicants seemed to have been able to attract more funding (including and excluding the funding received from Data Pitch) compared to applicants to Data Pitch that did not get funded. Furthermore, successful applicants seemed to have generated more employment than unsuccessful applicants. However, there is no evidence of stronger performance in terms of revenue growth. The comparison is based on a small sample of 9 unsuccessful applicants.

To gauge the net impact of Data Pitch, a comparison with unsuccessful applicants can be useful. This assumes that start-ups that applied to Data Pitch but were unsuccessful are similar to the successful applicants in at least some respects that are relevant for business success (e.g. formal qualifying criteria, interest in open innovation, motivation to participate).

As a complement to the successful applicants' survey, data was collected on unsuccessful applicants through a survey[49]. The survey reached in total 9 SMEs of which 8 are still active. The low number of responses from unsuccessful applicants limits the insights that can be obtained. For example, it is possible that more successful businesses had a greater motivation to respond, which would introduce an upward bias into the performance of the counterfactual group.

Nonetheless, the performance of the unsuccessful applicants can be compared with performance of the successful ones. The unsuccessful applicants survey asked about investment attracted, the change in revenue since applying to Data Pitch and the change in employment since applying. These were matched with data from the bi-weekly monitoring updates and 6 months progress update provided by successful applicants.[50]

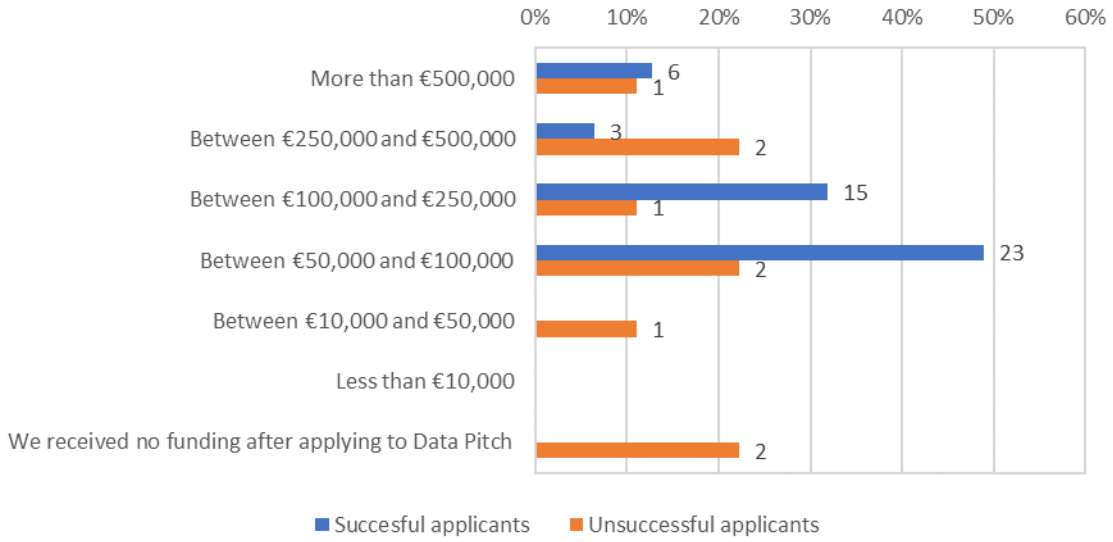
Successful applicants received, on average, more external funding (not including the €100,000 received through Data Pitch). Some applicants, both successful and unsuccessful, were able to attract substantial investments upwards of €500,000.

[49] See Annex A3.5

[50] The bi-weekly monitoring requirements cover approximately the 6 months during which the Data Pitch programme was run. The 6 months progress update captures successful applicants' progress in the 6 months following Data Pitch and is available for cohort 1 only at the time of writing. Where available, the data from the bi-weekly monitoring updates and the 6 months progress update were summed to arrive at a 12-month progress since the start of the programme. For cohort 2, the data from the bi-weekly monitoring updates was multiplied by 2 to provide comparable data.

The data on the unsuccessful applicants from the survey covers the 12 months since applying to the Data Pitch programme.

Figure 34 Funding received since applying to Data Pitch; % of respondents



For unsuccessful applicants: Data derived from Q13: How much external funding did you receive in the first 12 months after applying to Data Pitch? N = 9

For successful applicants: data derived from bi-weekly monitoring updates and, where available, the 6 months progress update. Funding received **includes** €100,000 funding received through Data Pitch.

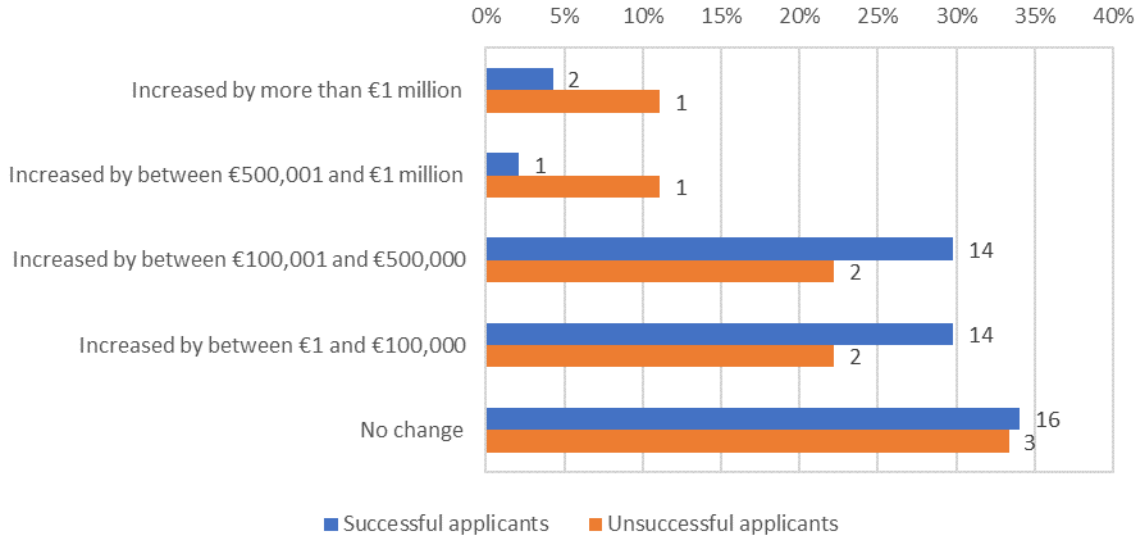
Exact number of responses in labels.

Source: Unsuccessful applicants survey, bi-weekly monitoring updates for cohort 1 and 2, and 6 months progress update for cohort 1

Concerning revenue gains, the successful applicants were not necessarily better off than the unsuccessful applicants. A slightly larger share of successful applicants reported no growth in revenues whereas a larger share of unsuccessful applicants reported increases upwards of €500,000.

However, the data derived from the monitoring updates and 6 months progress update may not account for the fact that Data Pitch could have fostered relationships between start-ups and a potential client; the data provider. Therefore, revenue effects of Data Pitch may only become apparent after some time.

Figure 35 Increase in revenue since applying to Data Pitch; % of respondents



For unsuccessful applicants: Data derived from Q11: By how much did your annualised revenue change in the first 12 months after applying to Data Pitch compared to before applying? N = 9

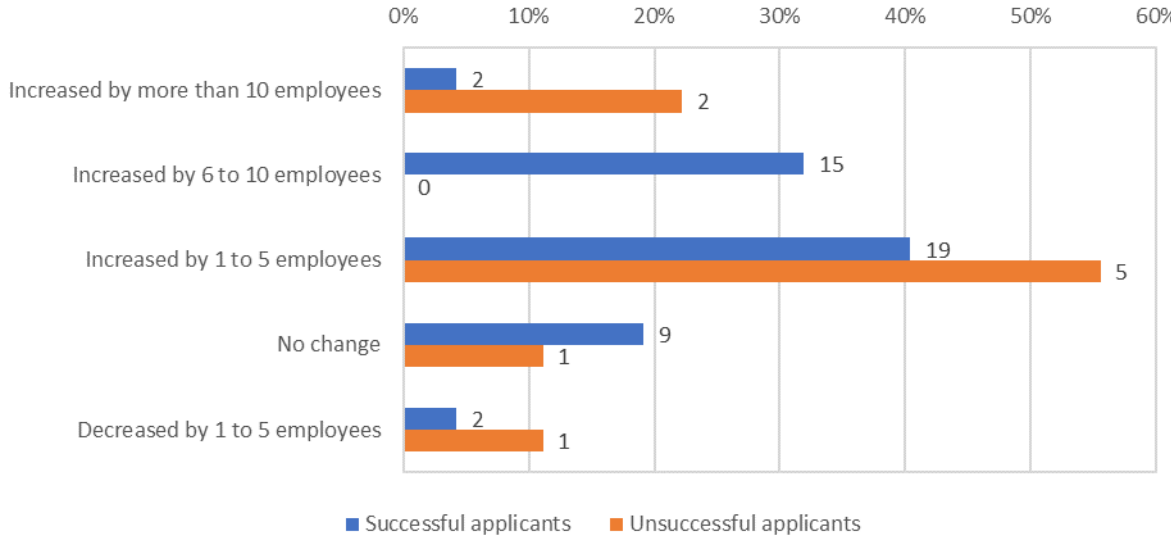
For successful applicants: data derived from bi-weekly monitoring updates and, where available, the 6 months progress. More precisely, this is derived from the sum of reported sales and efficiency gains. This assumes that the reported revenues in the monitoring updates capture only additional revenues directly related to the solution developed through Data Pitch and that reported efficiency gains capture the effect of the Data Pitch programme on other revenue streams.

Exact number of responses in labels.

Source: Unsuccessful applicants survey, bi-weekly monitoring updates for cohort 1 and 2, and 6 months progress update for cohort 1

Lastly, changes in employment can be compared. Successful applicants seemed to have fared better here. Generally, successful applicants have grown more quickly with fewer successful applicants reporting a decrease in employment and a substantial share showing an increase of 6 to 10 employees.

Figure 36 Change in employment since applying to Data Pitch; % of respondents



For unsuccessful applicants: Data derived from Q10: By how much has the number of employees in your company changed in the first 12 months after applying to Data Pitch? N = 9

For successful applicants: data derived from bi-weekly monitoring updates and, where available, the 6 months progress update.

Exact number of responses in labels.

Source: Unsuccessful applicants survey, bi-weekly monitoring updates for cohort 1 and 2, and 6 months progress update for cohort 1

The increase in employment for successful applicants is not surprising. The interviews with start-ups clearly showed that Data Pitch funding was used most often to hire people. The successful applicants survey further showed that Data Pitch funding enabled start-ups to acquire a broad range of skills, of which data science and machine learning, and software development were considered the most important.

The change in employment shown in Figure 36 forms the bedrock of the projection of employment growth in the 3 years following the programme, described in the next section.

3.9 Impact projections

Employment is forecasted to grow by, on average, 32% per year until 2022. This implies an increase of employment to 717 from 407 at the start of the programme. Forecasted employment growth is better than is predicted in a scenario without the existence of Data Pitch.

Annual total revenue generated by start-ups is forecasted to grow by 73% per year until 2022. Annual revenues are expected to by 459% of the total funds available to the Data Pitch consortium, including but not limited to the funding received by the start-ups in the programme.

This section presents estimates of potential future employment and revenue growth for the next three years based on a simulation model[51]. The details of the methodology are described in Annex 2.

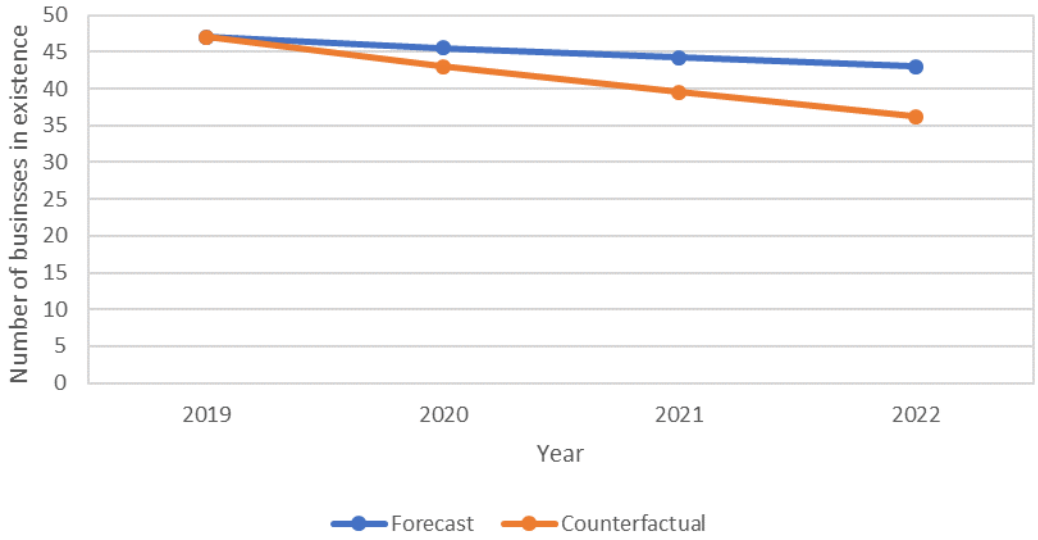
[51] This timeframe has been chosen because one of the input assumptions is explicitly valid over this time period.

3.9.1 Employment growth

The forecast model shows that Data Pitch may help start-ups grow more quickly. The figures below show the predicted number of funded start-ups that survive per year, and their average and total employment. Forecasts start in the year 2019 and go up to 2022.[52]

The forecast model shows that Data Pitch helps more start-ups survive the first three years after the programme. The projected number of surviving businesses in the forecast model is 43 by 2022, compared to 36 in the counterfactual model.

Figure 37 Survival rate in forecast and counterfactual model

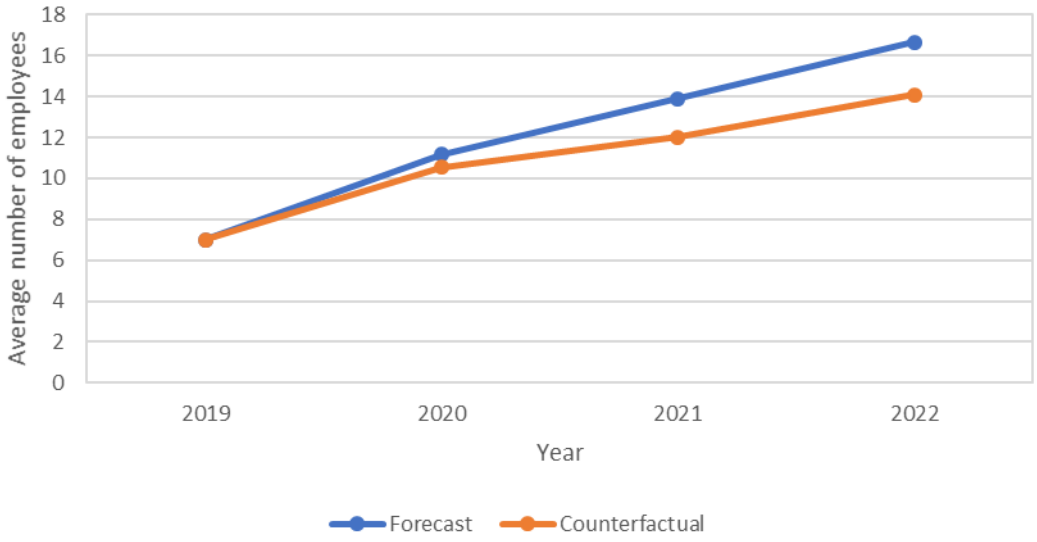


Source: Employment growth simulation

Similarly, Data Pitch stimulates the average number of employees for surviving start-ups. The average number of employees is higher in the forecast compared to the counterfactual without Data Pitch. With Data Pitch, the average start-up is expected to grow to 17 employees. This may only be 14 without Data Pitch.

[52] The model formally treats both cohorts as one, combined group and starts the projection at the start of the programme for this "combined cohort". For the sake of the projection, we equate the start of the "combined cohort" with the start of Cohort 2 in Spring 2019.

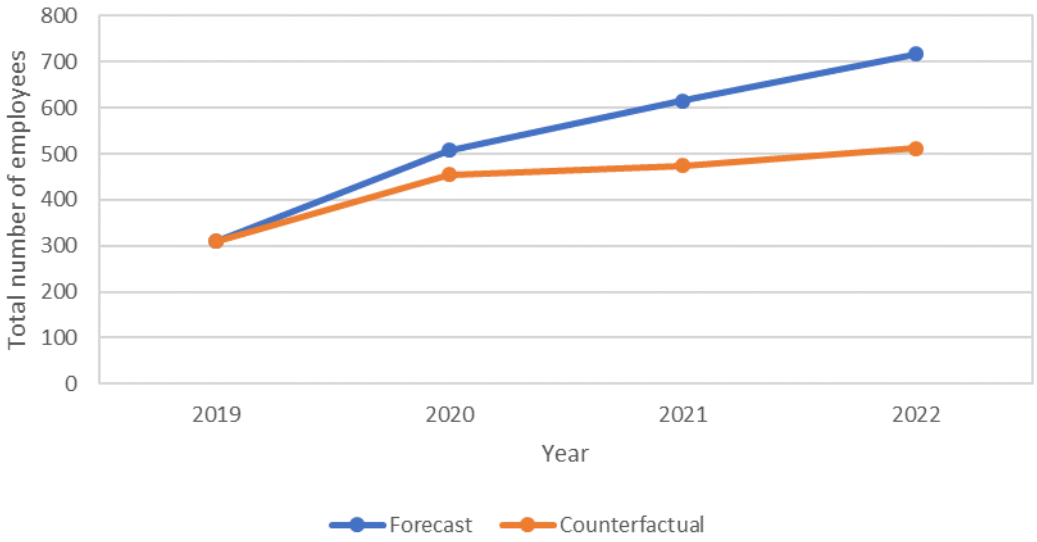
Figure 38 Average number of employees in forecast and counterfactual model



Source: Employment growth simulation

Given this, total employment is predicted to be substantially higher with Data Pitch funding. With Data Pitch, total employment is predicted to increase by 407 to 717 employees, or an annual growth of 32%. Without Data Pitch, the growth rate may only be 18% or a total increase of 202 employees.

Figure 39 Total employment in forecast and counterfactual model



Source: Employment growth simulation

3.9.2 Revenue growth

The simulation model estimating employment growth was also applied to potential revenue growth. Before discussing the results, important caveats must be addressed. The growth rates in the forecast projection are extrapolated from revenue generated during and shortly after the programme. Many start-ups in the programme were pre-revenue and therefore showed small or no growth during the programme. This carries through into the projection, potentially underestimating the true impact of Data Pitch.

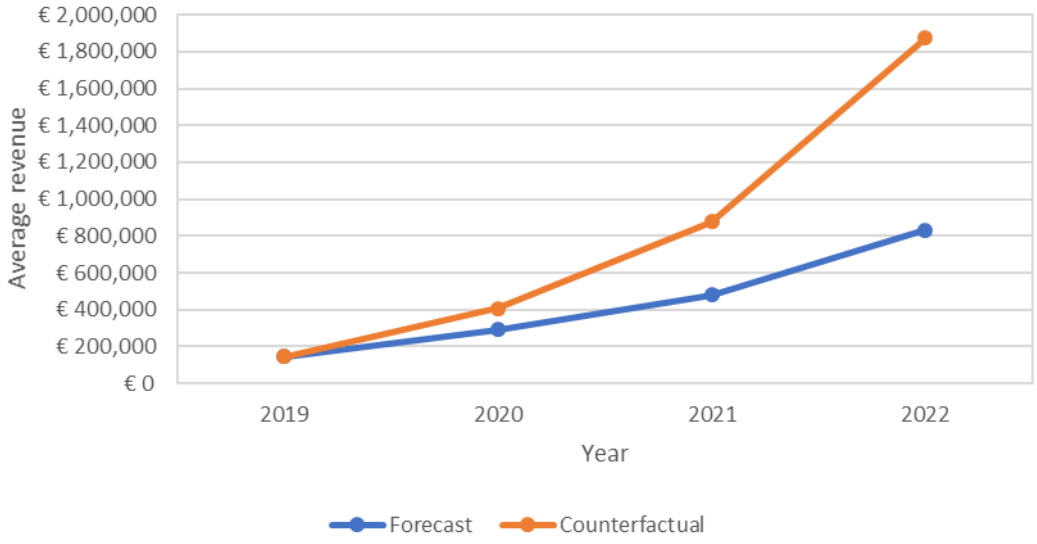
The growth rates in the counterfactual scenario are based on the performance of unsuccessful applicants. It is plausible that unsuccessful applicants that managed to generate stronger growth were more likely to answer the survey. Therefore, the assumptions in the counterfactual scenario may be biased in favour of higher growth. Furthermore, the counterfactual scenario is much less certain as it is based on substantially less data (9 observation).

This means that the forecast and counterfactual should be compared with caution. Moreover, the projected revenue growth is driven – in both scenarios – by exceptional performance of individual companies. These start-ups reported revenue increases of more than €1 million over one year, which is plausible for tech sector start-ups, but makes predictions of aggregate performance difficult.

In the end, the model assumptions in the projection model are grounded in reality as they are based on data reported by successful applicants during and after the programme and unsuccessful applicants in a survey. However, like any model, it hides complexity for the sake of tractability and the exact assumptions it uses will influence the results.

Considering these caveats, the figures below present average and total revenue projections for the period 2019-2022. The forecast model predicts that average revenue will grow from €147,723 in 2019 to €833,555 in 2022. Combined with the success/failure rate of businesses, this implies a growth of total annual revenue to €35,784,385 from €6,896,000. This is equivalent to a growth rate of 73% per year up until 2022. This growth rate may not be sustainable in the long run but shows that a decent proportion of start-ups funded by Data Pitch should become sustainable.

Figure 40 Average annual revenue in forecast and counterfactual model

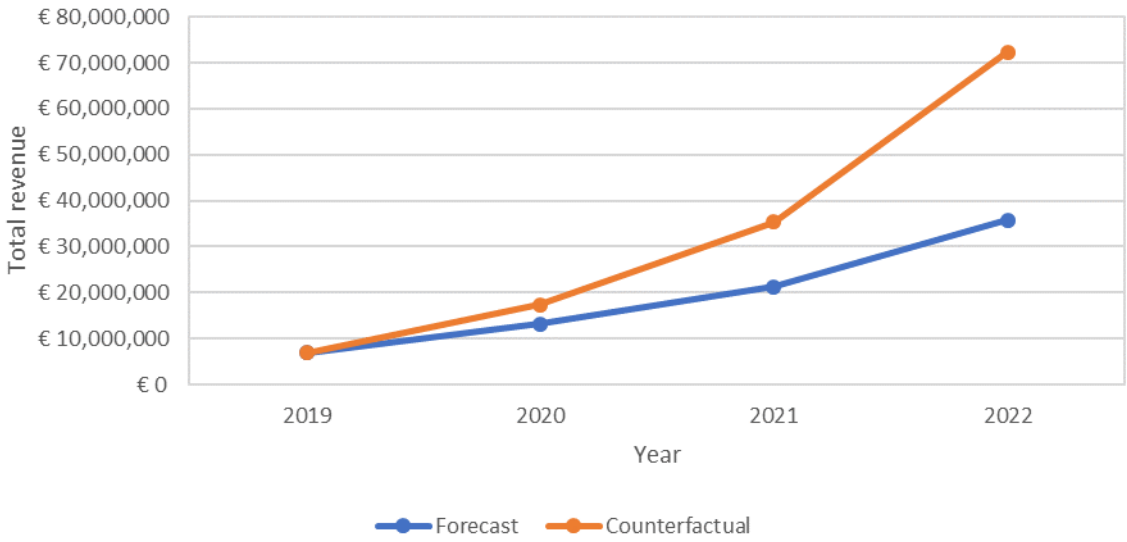


Source: Revenue growth simulation

In the counterfactual scenario, average revenue is expected to grow to €1,873,747 and total revenue to €72,374,032. This implies an annual growth of total revenue of 119%. The counterfactual scenario predicts higher growth than the forecast, which may show that Data Pitch has had less impact on revenue growth than on employment. However, as noted above, the projections may bias the counterfactual scenario up and the forecast scenario down. Considering these biases, the simulated performance difference between the two scenarios may be much less than shown here. This is exacerbated by the fact that pre-revenue funding for start-ups in Europe is widely available and

unsuccessful Data Pitch applicants had access to other sources of funding, including other accelerators.

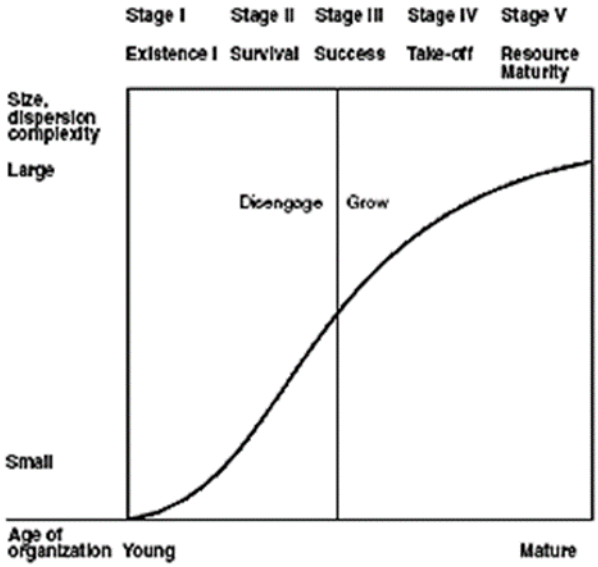
Figure 41 Total annual revenue in forecast and counterfactual model



Source: Revenue growth simulation

Overall, the simulation model provides a reasonable estimate of future impacts of Data Pitch participants in terms of employment and revenue growth, although the high annual revenue growth will not be sustainable in the long run. Small business *growth* naturally declines as businesses grow larger, as shown in Figure 42. Employment is predicted to grow with 32% whereas revenue is predicted to grow with 73% per year up until 2022.

Figure 42 Growth stages of small businesses



Source: Churchill & Lewis (1983)

Importantly, the focus on the Data Pitch start-ups does not consider the impact of Data Pitch on data providers and on the open innovation ecosystem in the future. Qualitative evidence strongly points to a positive impact on future participation in open innovation by participating data

providers, which is likely to multiply the benefit of the programme in a relatively short space of time. For some data providers, open innovation through the Data Pitch programme led to new services and products they can use in their operations. For others, the co-operation with start-ups led to an appreciation of the value of their data and the value of open innovation. Data Pitch generated a number of use cases that can be used as evidence for other potential data providers to engage with open innovation and "Data Pitch-like" data sharing arrangements.

3.9.3 Return on Investment 3 years after Data Pitch

Following Section 3.5.2, the predicted revenue generated by start-ups funded through Data Pitch can be set off against the total monetary resources made available to Data Pitch. The table below reproduces the Return on Investment (ROI) calculated in that section and extends this to include the ROI based on annual revenue by 2022 from the forecast projection.

Note that in the table, the ROI during the programme and respectively 6 and 12 months following the programme are cumulative as described in Section 3.5.2. Using cumulative sales is appropriate for the calculation of ROI since this captures all benefits that stem from the initial investment over time.

The ROI based on the projection, however, is not cumulative but based on predicted **annual** revenue by 2022. The annual revenues in the forecast projection have not been discounted. Therefore, the cumulative ROI derived from the forecast can be arbitrarily changed by increasing or decreasing the timespan over which it is calculated. This is avoided by providing an ROI based on annual revenue.

Table 12 Return on Investment; projected revenue

During the programme	6 months following the programme	12 months following the programme	Projected annual revenue by 2022
23%	91%	103%	459%

Note: data on 6 and 12 months after the programme are based on data from Cohort 1 only. The figures account for this by adjusting the investment provided through Data Pitch based on the number of start-ups for which data is available.

Source: Bi-weekly monitoring updates, 6 months progress update, 12 months progress update, revenue growth projection

As the table shows, revenue is expected to be 459% of the total resources available to Data Pitch by 2022. This shows that the investment in Data Pitch will likely produce viable businesses with substantial revenues.

3.10 Additionality

Taking into account factors that may decrease the net positive impact (programme 'additionality') of Data Pitch, we conclude that the impact of Data Pitch is likely to be strictly positive, especially once longer term impacts are taken into account: Data Pitch laid the groundwork for ongoing engagement in open innovation by some major European data providers, who appear ready to pursue their own initiatives going forward, and acted as a demonstrator for data-driven open innovation.

In any evaluation, a question that needs to be answered is how many of the benefits that can be attributed to the programme would have occurred anyway, and to what extent the project produces secondary consequences (both positive and negative)[53]. Relevant considerations are:

- **Deadweight:** how much of the impact would have been achieved anyway? Given the good availability of early-stage funding in Europe (confirmed by the fact that many of the Data Pitch participants had received external funding from other sources before they entered the programme), it is likely that high-quality start-ups like the ones selected for Data Pitch would have succeeded in some form anyway. Indeed, the survey of unsuccessful applicants suggests that some of the start-ups that were not selected achieved impressive revenue growth without the support from Data Pitch. However, in relation to the specific goal of fostering open innovation through data sharing, the evidence is strong that the participants would not have been able to access the data that Data Pitch enabled them to access. Moreover, there are few programmes that work with data providers as well as start-ups[54], it is unlikely that the benefits for data providers and the wider ecosystem benefits would have been achieved. In this regard, Data Pitch addressed a clear market failure, namely the frictions that inhibit mutually beneficial data sharing across organisations, which constitutes a net (strictly additional) benefit of the programme.
- **Leakage:** there is a possibility that some of the benefit of Data Pitch will 'leak' as supported start-ups move their activities overseas. One of the start-ups remarked in an interview that they were considering applying to Y Combinator, which would entail a relocation to the United States. However, for the vast majority of the organisations involved in Data Pitch, this is unlikely.
- **Displacement:** it is possible that Data Pitch displaced some investment by data providers. Some of the data providers were interested in open innovation before coming into contact with Data Pitch and might have pursued comparable initiatives under their own steam. However, it is clear that Data Pitch offered unique advantages in addressing the information-based market failures that are typical for the data economy, thereby getting projects off the ground that would not have happened otherwise.
- **Substitution:** for both the data providers and the start-ups, it is likely that some degree of substitution of other inputs for inputs from Data Pitch occurred. Some of the services provided as part of Data Pitch (legal advice, marketing support, GDPR training, etc.) would have been available from third-party providers. However, the identification of participant requirements and the tailor-made package provided by Data Pitch, based on the considerable experience of the consortium in the open innovation space, makes it likely that such third-party provision would have been less efficient and would not have been used to the same extent.
- **Long-term impacts** are likely to be strictly positive. By acting as a demonstrator for open innovation and involving new organisations in the practical implementation of open innovation projects that enabled data sharing, Data Pitch can be expected to generate indirect and long-term benefits. Specifically, Data Pitch laid the groundwork for ongoing engagement in open innovation by some major European data providers, who appear ready to pursue their own initiatives going forward.

[53] See HM Treasury (2018).

[54] A prominent example outside Data Pitch is the European Data Incubator (https://edincubator.eu/)

3.11 Evidence on the impact of start-up incubators & accelerators

The evidence on the impact of start-up incubators and accelerators is mixed. There is some evidence that incubators help start-ups grow. Furthermore, there is some evidence that accelerators help with generating, *inter alia*, investment, revenue and customer bases. The experience of Data Pitch is consistent with the general pattern of findings in the literature.

To put the performance of Data Pitch into perspective, this section summarises key findings of the wider literature on the impact of incubators and accelerators. The key sources is a study by innovation agency Nesta on behalf of the Department of Business, Energy and Industrial Strategy (Bone et al., 2019), which provides a literature review of previous research on accelerators and incubators across the world.

Establishing the impact of incubator and accelerator programmes is challenging. Firstly, there is no single accepted success metric by which programmes should be evaluated. Moreover, some metrics that seem to be related to success may have a more complex interpretation. One example is the failure rate of programme participants. High failure rates may not be bad. Non-viable businesses should be closed sooner rather than later. Good programmes may help entrepreneurs understand that certain businesses are not viable, leading to high failure rates that nonetheless have a positive impact. For programmes like Data Pitch, focusing on failure rates may risk leaving unviable firms on "life support", i.e. money injections into start-ups that should be allowed to fail.

Secondly, programmes typically select participants that are already more likely to succeed. This selection bias means that programme participants should do better than unsuccessful applicants, even if a programme would not provide any services. This selectiveness also creates a signal. Participation in a programme may signal to potential investors that a business is viable, because the business already passed an "inspection round". This signal may attract investors, even if the programme itself does not provide any services. In Data Pitch, especially the participants in sector challenges noticed the potential of signals. Being selected for the programme provided sector challenge participants with a signal that they could be trusted which strengthened their relationship with their data provider.

Notwithstanding these issues, some research on the impact of accelerators and incubators has been done. The results are mixed.

Research on incubators suggests that these types of programmes may help participants grow in employee size. However, the impact on survival and failure rates are mixed. These mixed results are difficult to interpret because, as noted above, both high and low failure rates could be a sign of success. Regarding innovation, little research has been done but incubators may not necessarily stimulate innovation as proxied by R&D intensity.

Evidence on accelerators is stronger but still mixed. There is some evidence that participation in accelerators may help businesses:

- attract investment;
- grow employment and revenue;
- reduce time to acquisition;
- close if they are unviable; and,
- gain customers.

The ability to help businesses gain customers may be particularly relevant for Data Pitch. The Data Pitch programme did not only provide funding and services, but also matched programme participants with potential customers, i.e. data providers.

However, the potential benefits noted above rarely hold for all accelerator programmes. For every accelerator that seems to help grow employment, there is another for which no impact has been found. Overall, research on general impacts of incubators and accelerators has been inconclusive.

Beyond the impact of the programmes themselves, there has been research into the impact of programme design. There is some evidence that support services that go beyond mere funding have a positive effect on the programme's performance. There is moderately strong evidence that programmes that provide networking opportunities and mentoring schemes are more successful. This shone through in the Data Pitch programme. Advisors and mentors[55] were generally well received and considered useful by participants interviewed for this assessment. But as before, there is no support service that is universally associated with success.

3.12 Comparison with ODINE and Nesta-reviewed accelerator

Data Pitch shares characteristics with other accelerators, some of which also focus on data.

Data Pitch compares most naturally with the ODINE accelerator. Data Pitch compares favourably with ODINE in terms of estimated future impact on employment.

This section provides a high-level comparison between Data Pitch, ODINE (IDC, 2017) and the corporate accelerator studied by Bone et al. 2019, for simplicity referred to as the 'Nesta accelerator'[56].

Table 13 Comparison of accelerator programmes

Accelerator	Funding	Mentoring and advisors	Training, workshops and/or webinars	(Shared) office space
Data Pitch	✓	✓	✓	
ODINE	✓	✓	✓	✓
Nesta accelerator		✓	✓	✓

Accelerator	Access to (investor) network	Peer-networking	Access to technology	Access to cloud and/or other services
Data Pitch	✓	✓	✓	✓
ODINE	✓		✓	✓
Nesta accelerator	✓		✓	✓

[55] In Data Pitch, the role of advisor and mentor were defined differently. The advisor came from within the Data Pitch consortium, tracked the participants throughout the programme and could provide general advice. Mentors were external to the consortium and were called in to provide more specific advice (e.g. technical or legal) at the request of participants. There seems to have been some confusion about the terminology regarding advisors and mentors among programme participants.

[56] Bone et al. (2019) originated from Nesta, hence the accelerator is indicated as the 'Nesta accelerator'. Note that the accelerator was not run by Nesta itself. The report does not name the organisation behind the accelerator.

Accelerator	Legal counsel regarding IP and data protection	Help with contractual arrangements	PR and communication
Data Pitch	✓	✓	✓
ODINE			✓
Nesta accelerator			

Accelerator	Virtual?	Centred around data?	Duration
Data Pitch	✓	✓	6 months
ODINE	✓	✓	6 months
Nesta accelerator			4 months

Source: Data Pitch documentation, IDC (2017), Bone et al. (2019)

Based on the forecast impact model in IDC (2017), companies funded by ODINE are expected to generate cumulative revenues of €110 million in the period 2016-2020 with 784 jobs generated. This corresponds to average revenue per company of around €1 million by 2020. ODINE achieves this by accelerating time to market, improving ideas and improving team skills. In a scenario where ODINE would not exist, the funded companies would have generated half as much cumulative revenue and 228 fewer jobs.

Companies funded through ODINE typically used multiple types of Open Data, with a particular focus on geospatial and environmental data. Furthermore, companies from countries with more mature Open Data markets were more likely to be successful in their application. Both point towards the value of fostering an environment conducive to Open Data.

The corporate accelerator reviewed by Bone et al. (2019) – headquartered in the US but with locations worldwide – was analysed using state-of-the-art and robust analysis methods[57]. Participation in the accelerator is associated with higher likelihood of survival, higher employment growth and more fundraising within five years from the application stage. The accelerator increased web presence – a proxy for survival – by 50%, helped accelerated businesses grow from 1-10 to 11-50 employees and increased fundraising by 77.6%[58] compared to similar businesses that were not accelerated.

These figures can be compared with the value generated by Data Pitch so far, and the forecasted value by 2022. As shown in section 3.5, both cohorts combined generated a total value of nearly €6.5 million[59] and 114 additional jobs while still in the programme. Including post-acceleration performance by cohort 1 (see section 3.5.1), confirmed value generated approximately equals €16.6 million[60] and additional employment of 140. The forecast model of section 3.9 predicts that employment generated by Data Pitch-funded start-ups will rise to 717, or an increase of 407, by 2022 and that annual revenue will grow with 72% per year until 2022.

[57] More precisely, Bone et al. (2019) used Fuzzy Regression Discontinuity Design.

[58] The average applicant to this accelerator raised about $150,000. Therefore, this percentage is equivalent to around £117,000.

[59] Sum of sales, investments and efficiencies as reported in the bi-weekly monitoring updates. Value generated excluding efficiencies nearly equals €5.7 million.

[60] Sum of sales, investment and efficiencies as reported in the bi-weekly monitoring updates, 6 months progress updates and 12 months progress updates. One cohort 1 participant reported additional investment in the range between €1 and €5 million. This has been included in this calculation as €2.5 million. Value generated excluding efficiencies is approximately €15.6 million.

Data Pitch compares most naturally with the ODINE accelerator. Data Pitch outperforms ODINE in employment generation. As noted above, Data Pitch-funded start-ups are predicted to generate additional employment equal to 407 by 2022. Over the same time span, the ODINE programme is predicted to generate 228 additional jobs (IDC, 2017, p. 47)[61] while funding more start-ups.

[61] In particular, IDC (2017) predicts that over a time span of three years, ODINE-funded start-ups will have generated 774 jobs from a base of 546. Hence, the increase equals 228. Note that IDC (2017) forecasts ODINE's impact for four years.

4 Case studies

These case studies were created from interviews with the chosen start-ups (and in some cases, the data providers), as well as data collected as part of the Data Pitch programme. These start-ups were chosen to present diverse examples of cases of how Data Pitch has promoted data-driven open innovation for companies of different sizes, maturities and sectors.

The solutions covered in the case studies range from supply-chain optimisation for industrial manufacturers, to traffic software which can predict safety and accident risk along a route. As well as giving a flavour of the breadth of innovative solutions instigated by Data Pitch, the case studies also provide an overview of the support received and overall benefits realised during the accelerator and afterwards, including the future outlook[62].

4.1 OBUU

Website: https://www.obuu.es/

Founded Year: 2015

Number/composition of staff on entry: 6

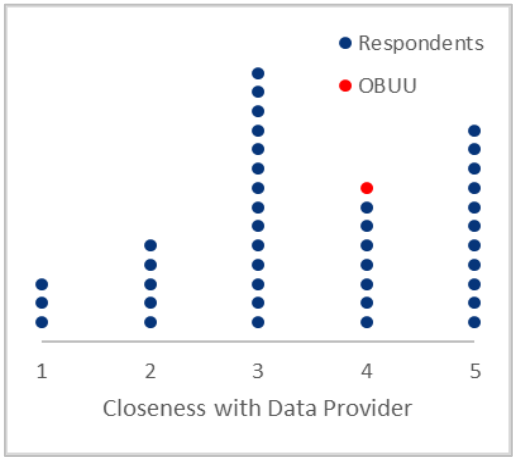
Challenge: DP3 - 2018 - Developing applications across manufacturing, logistics and supply chain with Greiner

Solution Tagline: StockWatch: optimisation of spare part stocks for industrial Machinery

[62] Two of the start-ups represented in the case studies (OBUU and Predina) have given permission to report detailed performance data. High-level summaries are provided in the other case studies.

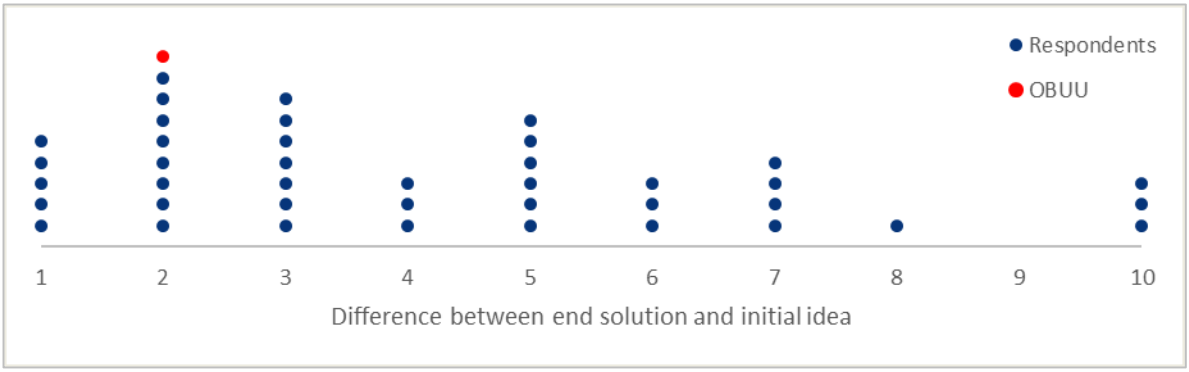
4.1.1 Solution[63]

As part of the data pitch programme, OBUU continued to develop a software solution they had been working on since 2015 (referred to as 'StockWatch'). This solution is an online, cloud-based platform that enables users to map out shipping, storage and manufacturing flows. The software then analyses these flows with input data to find where further efficiencies can be attained.



As part of their work with Greiner, OBUU utilised data from five of Greiner's production plants to determine the risk of spare parts not being available when needed (Stock Risk of Shortage; ROS), the average time the system is down when a part is not available (Mean Down Time; MDT), and the overall investment in the stock available. These three factors served as key performance indicators for determining the efficiency of Greiner's supply chain. Using StockWatch, OBUU found that Greiner could achieve a reduction of fixed asset investment of "at least a double-digit percentage".

StockWatch is currently being licensed by OBUU on a SaaS (software as a service) basis. As part of the Data Pitch Programme, Greiner obtained a one-year license. OBUU have also licensed this updated software to further clients in industries such as energy and aerospace.



4.1.2 Data

OBUU received 10 megabytes of data from Greiner; OBUU received a full copy of this data and therefore had full control over how the data was used and handled. The data contained information from one of Greiner's production plants on:

- Logs and reports of failure cases for Greiner's manufacturing machines;
- Logistics orders;
- Maintenance and usage history;

Number of Datasets used: 12

Update frequency of the data: Occasional/Irregular

Level of control over the data: Full

[63] The charts in these case studies display responses to a survey completed by Data Pitch participants. 41 out of the total 47 firms responded, the x axis refers to the answers given by each firm. One dot represents one firm's response to this question.

As Greiner were near the beginning of a new digitalisation phase, this data was collected and processed for the first time before being delivered to OBUU.

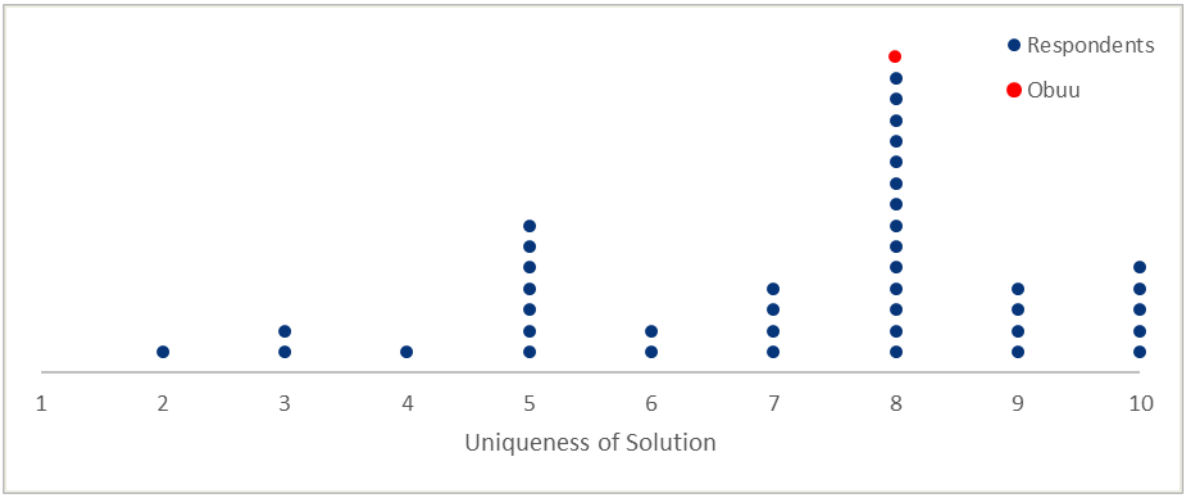
Size of the data: 0.01GB, 4000 observations

However, although the dataset contained the full set of variables OBUU were looking for, both OBUU and Greiner remarked that the data contained fewer observations than initially predicted (4,000). As a result, OBUU changed their initial strategy of testing one scenario thoroughly, to the creation and testing of four separate and smaller scenarios instead.

4.1.3 Benefits from the programme

Access to a data provider was an important part of OBUU's motivation for participating in Data Pitch. As OBUU already had a solution they had been developing, they saw Data Pitch as an opportunity to work with a large corporate client such as Greiner. They state that the data they received probably would not have been easily accessible outside Data Pitch. OBUU have also commented on how pleased they were with their interactions with Greiner, especially with regards to the active role their point of contact was taking; an example of this active role was their contact proposing possible solutions and routes their project could undergo.

For OBUU, a key benefit of Data Pitch was that they were given a ready-made use case of direct relevance to an industry customer. Additionally, this close collaboration with Greiner enabled OBUU to enter a new industry which will potentially enable them to sign contracts with other firms in adjacent industries.



OBUU also received just under €100,000 in funding from Data Pitch. Over 80% of this funding was spent on paying the salaries for their seven employees, with the bulk of the remaining funds being spent on travelling for meetings with Greiner, Data Pitch events and meetings, and attending events.

Resources Data Pitch funding enabled OBUU to acquire: Software development skills, Data science skills, Domain knowledge, Data analysis skills & Business Management skills.

Would you have been able to access the same data without Data Pitch? Yes, already had connections with data providers.

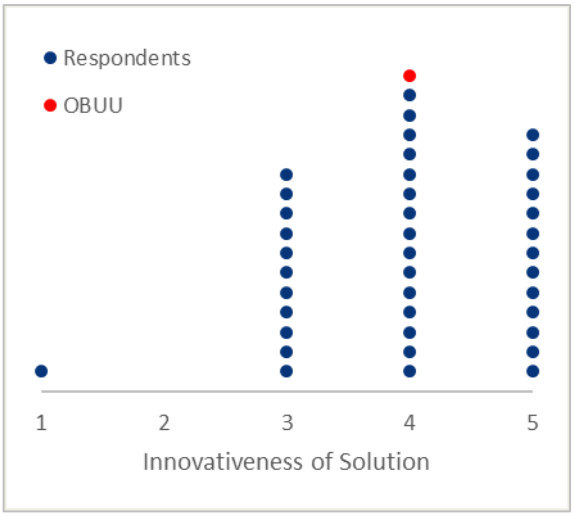
Other resources and benefits gained from Data Pitch included access to mentors, webinars & events. Whilst OBUU tried to attend the majority of these events, they appreciated that they did not have to attend events which were not relevant for their current business stage. The provision of Amazon Web Service credits from Data Pitch also enabled OBUU to host their StockWatch solution for the next couple of years without cost.

At the start of the programme, OBUU described successful participation with the following goals:

- To test their algorithms within a mass production environment and acquire an early adopter of their updated software;
- To acquire a new client and diversify into a new industry;
- To attend conferences and events organised by Data Pitch, where they will be able to make new contacts and showcase their projects to agents from European Industries.

OBUU have expanded their software and tested it using Greiner's manufacturing and logistics data, gained a number of new clients as well as attended events such as Maintenance Next in Rotterdam and the Paris Air Show; they also planned to attend the Packaging Innovations convention in November 2019. As a result, OBUU achieved the goals they set out at the start of the programme and attained several benefits as a result.

4.1.4 Acceleration Performance



During the accelerator period, OBUU made a gross profit of almost €40,000, part of which came from closing a contract with a client in the railway industry. At time of application, OBUU had accumulated a lifetime revenue of approximately €700,000 and during their six months in Data Pitch, they made €224,230 in revenue.

Additionally, OBUU have received €60,000 in investment alongside the amount received from Data Pitch. This investment took the form of funding for an R&D project into new statistical algorithms. At the time of application to Data Pitch, OBUU had raised a total of €602,000. In addition to this, they increased their employee count from six on entry (hiring one just before the project began) to nine on exit.

Revenue (€)	Gross Profit (€)	Investment	Realised Efficiencies	Employee increase	Increase in paying users
224,230.5	39,901.76	60,000	15,000	2 (7 –> 9)	1 (3 –> 4)

Note: These metrics track progress throughout the 6-month accelerator period.
Source: Bi-weekly monitoring updates and OBUU milestone reviews

4.1.5 Since the programme / Outlook[64]

OBUU have taken part in several similar programmes such as the Airbus Bizlab, ESA Business Incubation Centre and the Renfe TrenLab start-up incubator. Moving forward, OBUU are choosing to ramp-down their participation in accelerators and instead focus on growing in the industries they are now established in. As a result, they are planning to be 'pickier and more selective'.

In line with these plans, OBUU are focused on retaining their current customers, and creating recurrent demand for their services. They plan to do this by launching a Freemium campaign to showcase their product, as well as continuing the technical development of their product. This development includes improvements to user interface and ability to deal with a higher complexity of solutions.

Effort required for scalability:

Core solution – Minimal

Future adaptations of core solution - Medium

4.1.6 Insights

OBUU has been successful even though their initial application for the open innovation challenge was rejected. This suggests that success depends on achieving the right match between data providers (or challenges) and start-ups. Additionally, the increase in the number of matched companies in the second cohort may suggest that the matching process (and experience of those matching) had improved. Future programmes should consider the right mix of data providers and start-ups, and possibly increase the number of data providers relative to the number of start-ups to increase the probability of successful matches.

[64] In the survey, Core solution was defined as 'Markets & customers as currently identified'; Future adaptations of core solution was defined as 'new markets/application areas/customer groups'.

The data provider was matched with 5 start-ups and reflected that this may have been too many. For large companies (such as Greiner), working with start-ups on open-innovation programmes may represent a new challenge. Therefore, they may not fully be prepared to provide the support needed to help maximise the impact which these start-ups can provide through close collaboration. OBUU and Greiner have both stated that their relationship benefitted from the enthusiasm present from both parties. Future success in similar open innovation projects may be influenced by a high level of effort, enthusiasm and support from all participating parties.

Data Pitch worked efficiently in facilitating the match between OBUU and Greiner and was seen as providing a useful and comprehensive service, with minimum burden on the start-up. OBUU appreciated that there was no obligation to take part in parts of the programme that were not relevant for them. Apart from data access and collaboration with the data provider, funding (mostly for additional staff), cloud space to host the solution and some legal advice were the most useful elements of Data Pitch.

Open Innovation programmes function in an ecosystem of funding opportunities that include other accelerators, both corporate and publicly funded. Companies like OBUU often receive funding from different sources. Attributing start-up success to any one of these is difficult, and a more holistic view should consider interactions between different types of support and their cumulative effect.

The example of OBUU illustrates the importance of 'introductions' for start-ups to enter a new market or industry. This reflects the fact that challenges faced by business are often similar across sectors (e.g. supply chain optimisation), but that barriers continue to exist that limit the portability of solutions. The 'openness' required for open innovation to work is more than openness in terms of data access, it also includes domain, market knowledge, trust and openness to engage with entities whose track record is in different sectors. However, Data Pitch also clearly enabled the start-up to access data that would have been inaccessible otherwise.

The example also shows the importance of buy-in on the part of the data provider, both in terms of an institutional commitment to open innovation, and in terms of motivated and skilled personnel driving the cooperation with the start-up. The data provider had previous experience with open innovation and R&D cooperation more generally. Although the data provider had previous experience, their collaboration with OBUU provided a use case for their data which they had not previously considered. Greiner did not think they needed OBUU's product at the start of the project but ended up using StockWatch as part of their business operations. Data Pitch has therefore provided an example as to how a data provider's assumptions regarding the usefulness of their data can be challenged through collaboration and open innovation.

In many ways, the OBUU-Greiner match is an ideal example of the open innovation model as implemented by Data Pitch, with real learning on both sides and a useable product at the end. The start-up embraced the open innovation model exemplified by Data Pitch and intends to use it to enter other markets. However, it is noteworthy that OBUU has been operating for a number of years, participated in other accelerators and had a working commercial product going into the challenge. This may suggest that they were well positioned to conduct an effective and successful project with their data provider, as they did not have to contend with early-stage concerns and constraints (such as staffing and business direction).

4.2 Bliq

Website: www.bliq.ai

Founded Year: 2018

Number/composition of staff on entry: 14

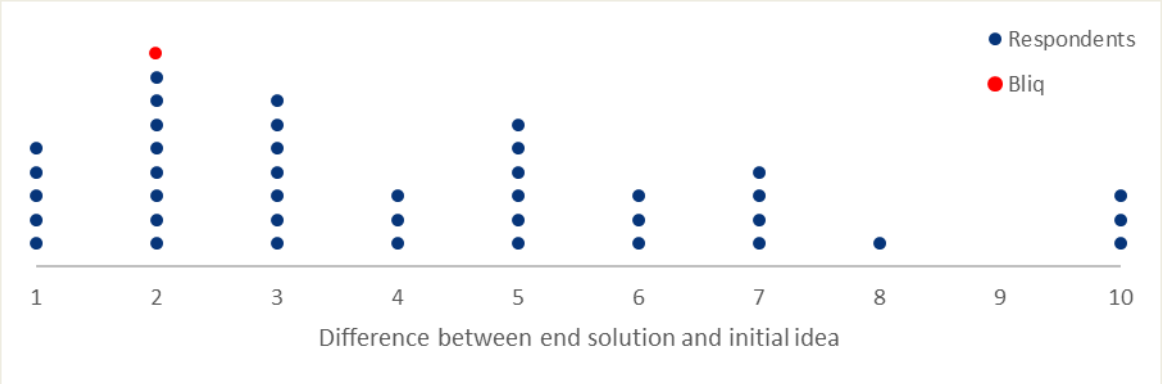
Challenge: SC7 - 2018 - Innovative solutions to improve mobility and reduce traffic congestion

Solution Tagline: Live parking maps for developers in mobility

4.2.1 Solution

Bliq, formerly known as AIPARK, create live parking maps. Bliq provide these maps and APIs for developers, who can then develop solutions that show downstream users where to find parking spots.

Before Data Pitch, parking spaces provided on these maps would have to be manually drawn and entered by Bliq. The solution they developed as part of Data Pitch enabled them to automate this process using machine learning.



As an example of the current and potential user base, Bliq's solution can be built into car entertainment systems to show the driver the location of available parking spaces near them.

4.2.2 Data

Bliq used over 500GB of data in the construction of their solution. This data, which they received from their data provider, took the form of high-quality satellite images. Bliq noted that although the quality of this data was very high, the update frequency was relatively low, at 12 – 18 months. Bliq did not have access to a full copy of this data, so they only had partial control over this dataset.

Number of Datasets used: 6

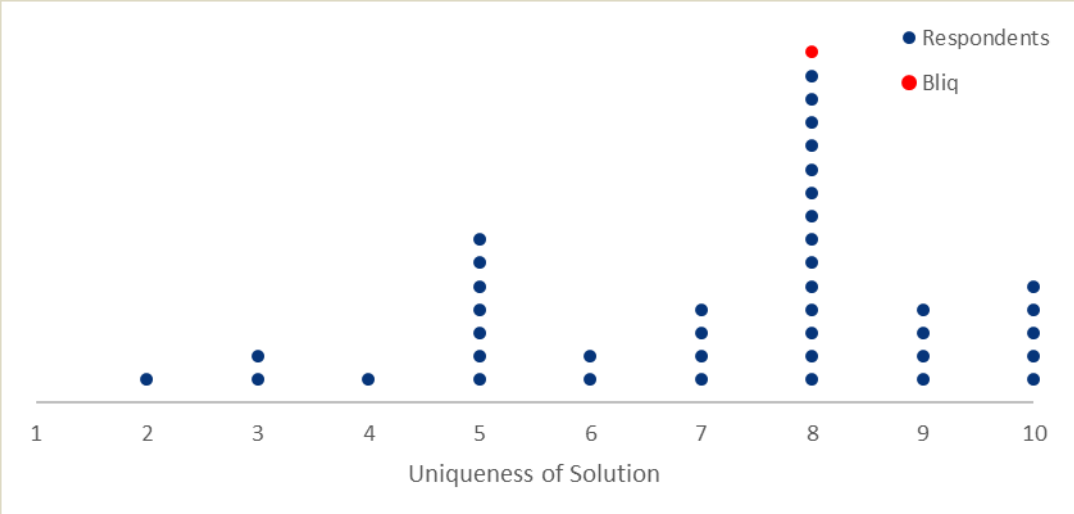
Update frequency of the data: Occasional/Irregular

Level of control over the data: Partial

Size of the data: 600GB, 20,200 observations

Bliq would have been able to access this data without Data Pitch, but their solution would have been much less mature. Bliq acquired this data through a revenue-sharing agreement with their data provider, who provided the data free of charge, as Bliq were investigating a use case Hexagon had not previously considered.

4.2.3 Benefits from the programme



The majority of the funding, received from Data Pitch, was spent on staffing costs, including the hiring of 2 additional engineers.

Resources Data Pitch funding enabled Bliq to acquire: Subject matter/domain knowledge, Software development skills, Data science/ machine learning skills

Would you have been able to access the same data without Data Pitch? Yes, but the developed solution would be less mature.

Data Pitch funding enabled Bliq to pursue a line of development which they would have been unable to do otherwise. This funding allowed Bliq to modify their development roadmap to include additional functionality. Bliq hopes to obtain additional input data from other providers in exchange for access to their software in the future.

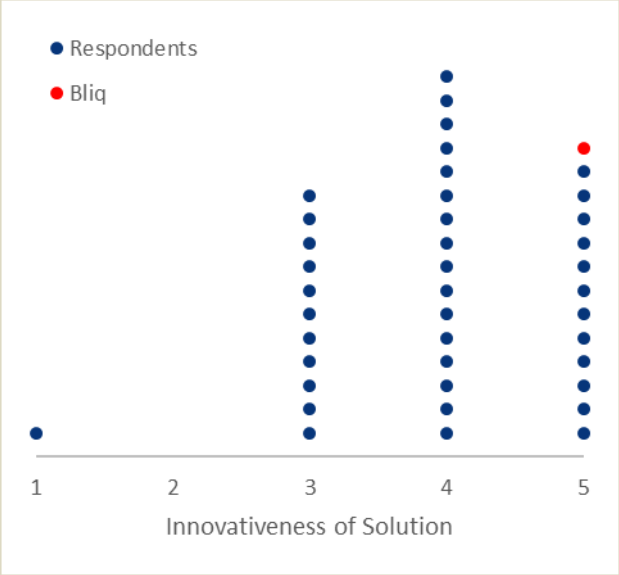
The bi-weekly check-in Bliq had with their advisor also provided Bliq with on-going feedback regarding actions they could take for improvement and next steps. Bliq comment that this access to an advisor, alongside the access to the network of the Data Pitch consortium, has been a benefit for them. The network has provided an opportunity for Bliq to contact new potential partners and clients.

4.2.4 Acceleration Performance[65]

During the accelerator period, Bliq generated an increase in revenue. Additionally, they also gained an increase in investment and realised efficiencies. At time of application, Bliq had already raised a significant amount in investment.

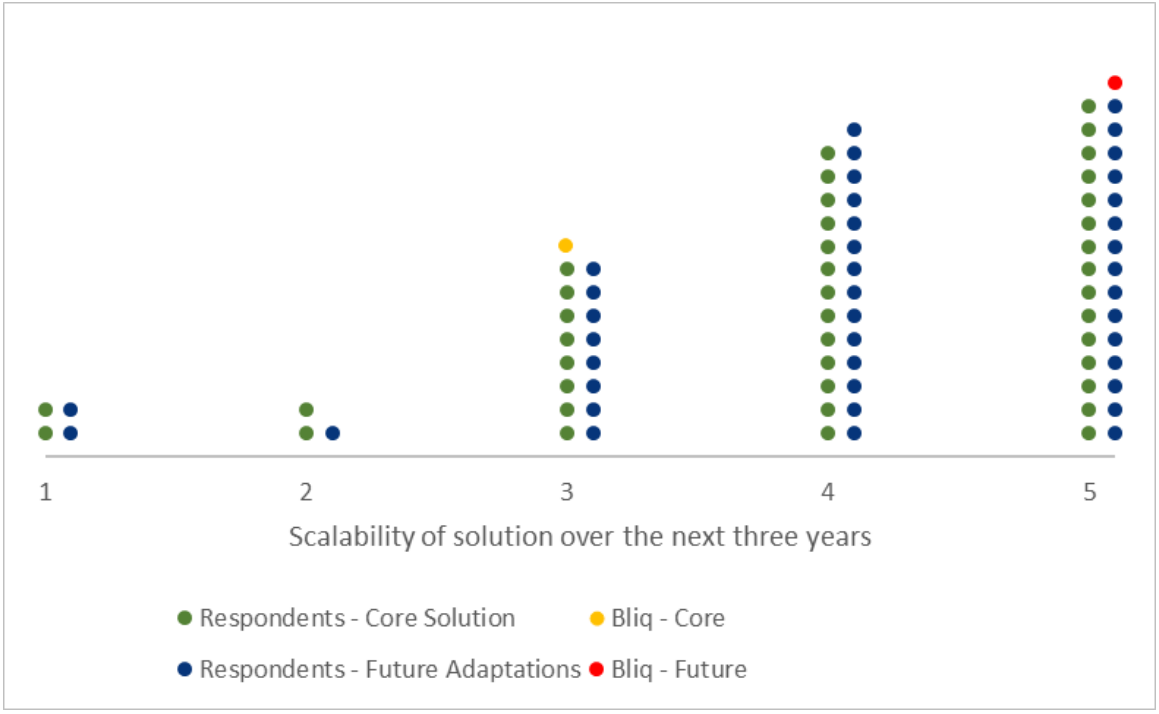
Throughout the accelerator period, their monthly userbase increased as well as their employee count, with the addition of new engineers.

Bliq also closed a deal with a Japan-based automotive supplier and are considering future joint product-development.



[65] Approval has not been given by Bliq to report precise figures

4.2.5 Since the programme / Outlook



Bliq's next steps revolve around market penetration. They are focusing on entering the USA market, as well as expanding their market presence in multiple other European countries.

Bliq are also continuing work on their mobile apps, with the launch of an early access programme in Berlin for their Mobile Parking Vision app. Bliq are also planning to develop a ride-hailing/taxi driving app to build proprietary data acquisition, as well as serving as a potential side business model.

Effort required for scalability:

Core solution – Medium

Future adaptations of core solution - Medium

4.2.6 Insights

Bliq is an example of a start-up for which the funding was the most important benefit of their participation in Data Pitch. The funding provided for Bliq enabled them to develop aspects of their solution which they had not previously planned to create. This product, which would not have been created without Data Pitch, would also help with attracting future investment. This access to additional investment can result in Data Pitch having a longer-term effect on business development beyond the immediate benefits from the accelerator.

The example also illustrates that start-ups can gain access to external data providers without help from third parties, in this case using a revenue-sharing arrangement with an upstream data provider. Bliq also provides a good example of AI-enabled efficiency gains (replacing a largely manual process with an automated one based on the exploitation of a rich set of training data).

4.3 Energeo

Website: www.energeo.co.uk

Founded Year: 2015

Number/composition of staff on entry: 3

Challenge: SC3-2018 - Increasing efficient energy creation and use

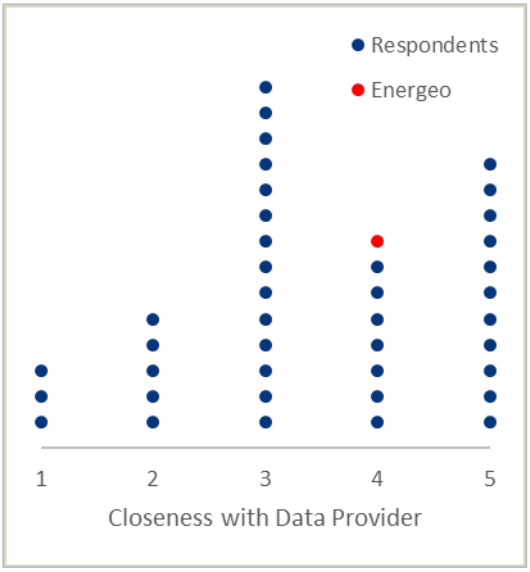
Solution Tagline: Energeo deliver Sustainable Intelligence for the low carbon economy

4.3.1 Solution

Prior to the programme, Energeo had developed a beta Software as a Service (SaaS) platform which:

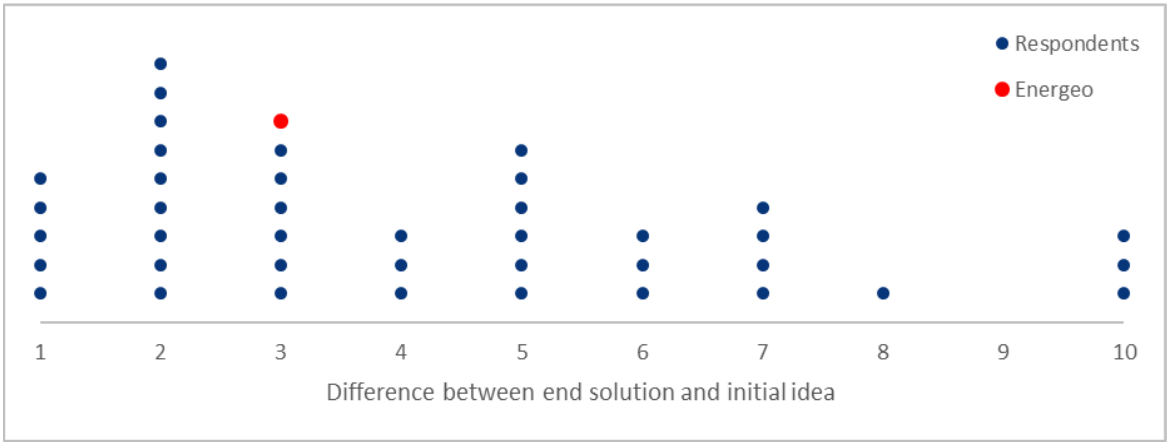
- Delivered intelligence created from a variety of geospatial inputs (such as OS maps, LiDAR, satellite imagery etc.)
- Integrated proprietary data to provide the end-user with a data analysis toolkit to support the identification of potential energy projects.

As an example, this solution could use satellite imagery to determine which rooftops were ideal for solar panels. This solution was aimed at local authorities and councils who would have the goal of reducing carbon emissions.



Their Data Pitch project was to continue development on this existing platform by integrating energy data from their data provider. This updated solution would be able to determine the effect that

additional solar panels would have on substations (e.g. the effect that excess solar energy would have if it would have to flow back into the grid).



4.3.2 Data

Energeo used both closed and open data sources as part of their solution. They had a previous relationship with their data provider, who had provided them a dataset prior to entering Data Pitch. This dataset contained near real-time data for substation/transformers located in their data provider's service area. This information included details on transformer temperature, ambient and inside temperatures, transformer loads and currents.

Number of Datasets used: 5

Update frequency of the data: Live

Level of control over the data: Partial

Energeo also received data on historic energy usage from a London University. This dataset was used to analyse trends and patterns of energy usage over a 12 month period. Energeo had partial control over these datasets as this data would be obtained via an API call when required.

Size of the data provided by Western power Distribution: 100 observations, 1GB

The additional datasets used include address data, satellite imagery and energy performance certificates.

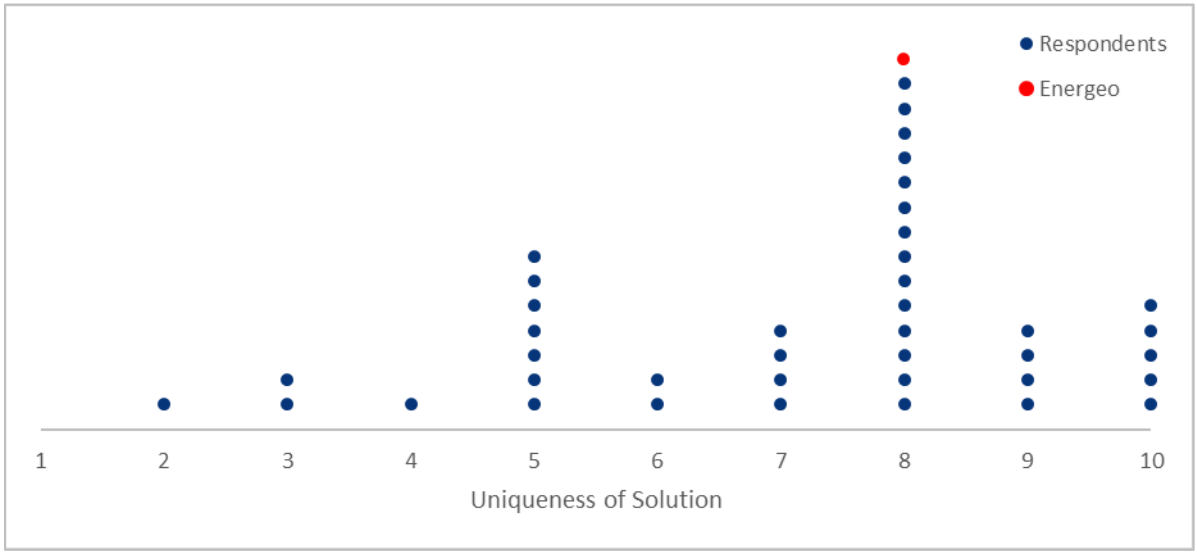
4.3.3 Benefits from the programme

At the start of the programme, Energeo had 1 full-time employee, and 2 part-time employees. They relied a lot on outsourcing and subcontracting to complete their previous contracts. With the funding they received from Data Pitch, they managed to hire additional employees. This enabled them to bring software development and geospatial resources into their company on a full-time basis.

Resources Data Pitch funding enabled Energeo to acquire: Subject matter/domain knowledge, Software development skills, business management skills, marketing/sales skills

Would you have been able to access the same data without Data Pitch? Yes, data provider gave access to this data to look at before Data Pitch, but Data Pitch served as the catalyst to bring these ideas to fruition.

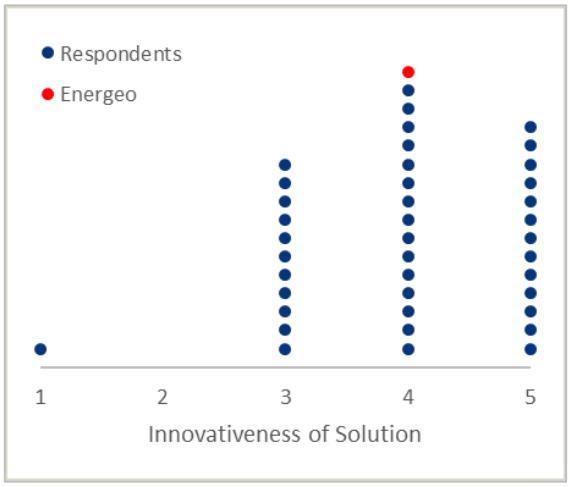
The Data Pitch programme enabled Energeo to progress along their planned direction much more rapidly. A large part of this was due to the creation of an MVP (minimum viable product) which was much more developed than they would have been able to do without Data Pitch.



From their data provider, Energeo received technical support with interpreting and utilising the provided electrical and energy-related data (as this was outside their geospatial expertise). They comment they did not receive too much support from Data Pitch (or asked for it), as they knew what they wanted to achieve from the programme and set upon achieving this.

4.3.4 Acceleration Performance[66]

During the accelerator period, Energeo generated an increase in revenue. They also increased their employment by more than double over the course of the accelerator period.

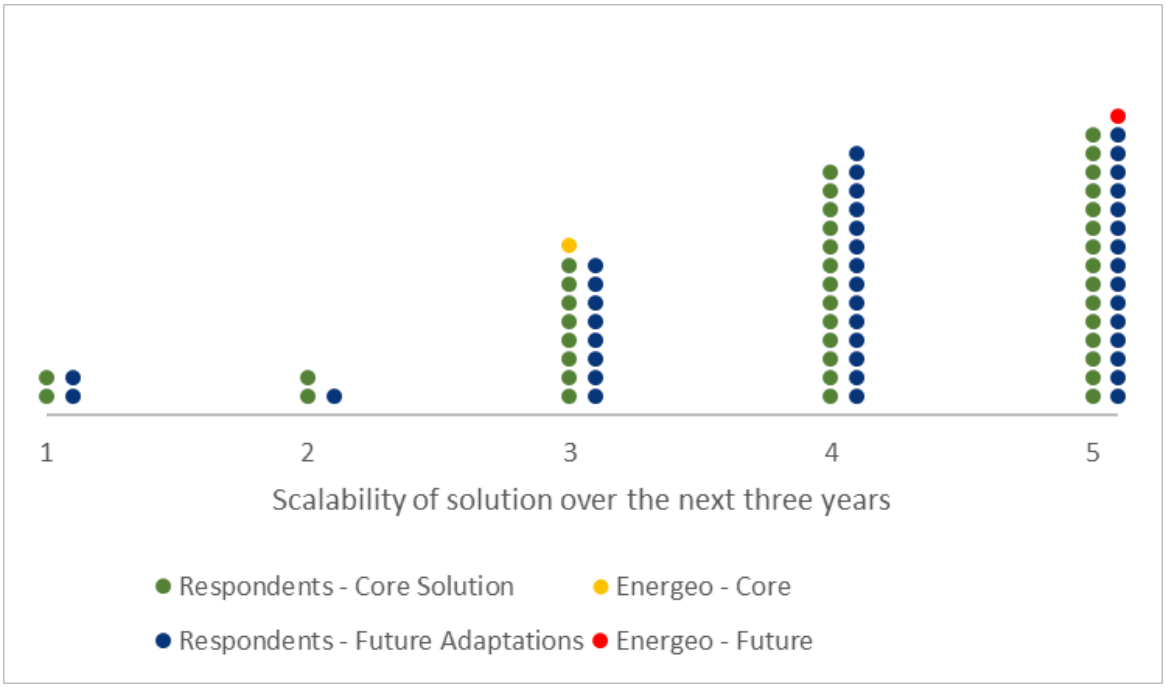


4.3.5 Since the programme / Outlook

By the end of the programme, Energeo had been accepted into the PwC 'Scale' GovTech accelerator programme, which began in October 2019. They have also obtained a contract with a British county council to provide geospatial intelligence regarding low carbon tech potential across the county.

[66] Approval has not been given by Energeo to report precise figures

They are completing this work as part of a consortium, in which Energeo's element could last up to 3 years.



Additionally, Energeo were named the most innovative sustainable energy products and service provider, and leading provider of EV charging infrastructure solutions in the SME news, 'Energy and Power Awards 2019'.

Additional contracts and business Energeo have secured include an order from an additional British county council for their solution as well as ongoing dialogue with potential resellers of their data services in South Africa and Ireland.

Effort required for scalability:

Core solution – Medium

Future adaptations of core solution - High

Energeo have a long-term ambition to turn the software they created during Data Pitch into a subscription-based platform; they have begun initial research into the possibility of attaining this.

4.3.6 Insights

The Energeo solution is one of the relatively few in Data Pitch that combine data from a number of different sources to create a new solution ('recombinant innovation'). Energeo used the funding received from Data Pitch to accelerate along a well-defined development path, enabled mainly by hiring additional full-time resources. Technical and market knowledge supplied by Data Pitch were also important for the start-up. Data Pitch appears to have functioned like a more traditional accelerator. It highlights the interplay between innovative start-ups, different data providers and the support ecosystem (including other accelerators).

Although Energeo was able to access the data used without Data Pitch, it was the participation in the programme which was the catalyst for Energeo to generate impact from the data. This is an example of how programmes such as Data Pitch can still encourage innovation through the opening of data, even in cases where data was accessible for start-ups beforehand.

Energeo did raise a suggestion regarding the payment structure of the programme. They commented that for a small, or young, start-up, cashflow is critical, and that the payment structure could be modified to reflect the needs of these smaller SMEs. Specifically, the gap between the first and second payments made it difficult for Energeo to manage their project on this payment timeline.

4.4 Predina

Website: https://www.predina.com/

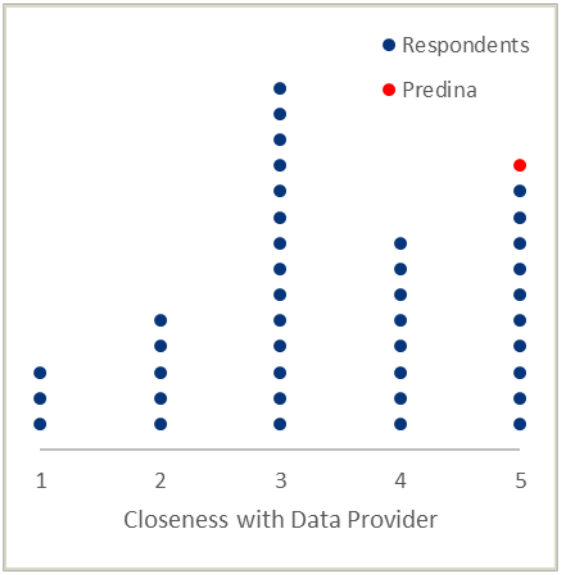
Founded Year: 2016

Number/composition of staff on entry: 5

Challenge: SC5-2017: How can we use data to make manufacturing, logistics and maintenance processes more efficient and able to support new models of use and repair

Solution Tagline: Guardian, AI technology software that predicts the risk of road accidents dynamically and accurately in real time: "Google maps for safety"

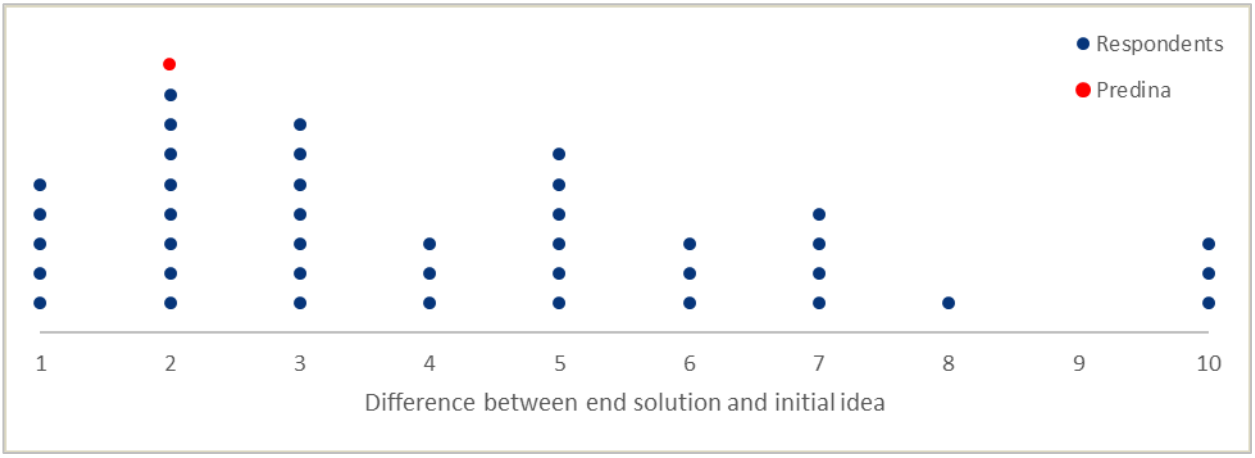
4.4.1 Solution



Predina developed two solutions as part of Data Pitch, which fed into a final product. The first of these solutions was an optimised algorithm which was used to predict the risk of road accidents. The second of these solutions was a data layer which would connect to a sat nav product; this product would then be able to alert drivers to the risk of road accidents. Both solutions came together in a web-based tool which enabled users to plan their routes, along with the expected safety of their route. The collective name for this software package was 'Guardian'.

These solutions were trained and developed using driver data from their data provider. This driver data described the actions and characteristics of truck drivers in their data provider's delivery fleets.

During Data Pitch, Predina's targeted client base had evolved to the provision of systems directly to car manufacturers, and the provision of systems for the automotive aftermarket. The aim was to use these systems for the purposes of insurance analytics and fleet tracking.



4.4.2 Data

Predina obtained 2 million observations, totalling 30GB in size, of data from their data provider. This dataset contained information on:

Number of Datasets used: 2

Update frequency of the data: Occasional/Irregular

Level of control over the data: Full

- Comprehensive accidents data for all countries (countries which their data provider operate in): driver statement, images, videos of before, during and after time of accident, affected parties, compensation costs etc.;
- Near miss data for all countries;
- Comprehensive driver behaviour data: this included profiles for each driver, their respective depots, previous accidents they have experienced at work, heavy acceleration and quick break events, amongst other similar indicators;
- Real-time driver location tracking through a TomTom API;
- Database of schedules generated each day for all drivers by the client's scheduling team.

In addition to this, Predina also used several public datasets (both open and paid). These datasets contained data on weather, police records, traffic volumes, fuel prices and location-specific mapping data. Predina had full control over the data received from their data provider, as well as these public datasets they used for their solution.

Size of the data received from their data provider: 30GB, 2 million observations

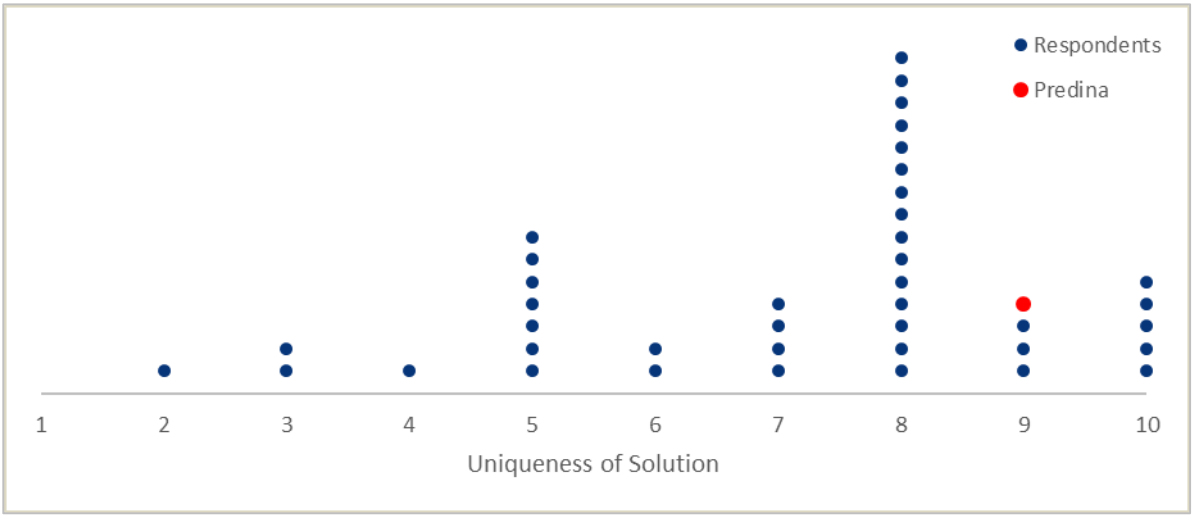
4.4.3 Benefits from the programme

The Data Pitch funding was mostly spent on staffing and salary costs, with the hiring of additional engineers. They also used the funding for Amazon Web Services and Google Cloud credits, which were used to host their product.

From Data Pitch, Predina have commented that the ability to network with other start-up founders was a 'very vital' benefit for them. The provision of an environment where Predina could meet other founders, as well as share problems and solutions with each other, was a substantial benefit of the programme.

Resources Data Pitch funding enabled Predina to acquire: Subject matter/ domain knowledge, software development skills, ICT infrastructure/hardware, Other IT skills, ICT infrastructure/hardware, marketing/sales skills

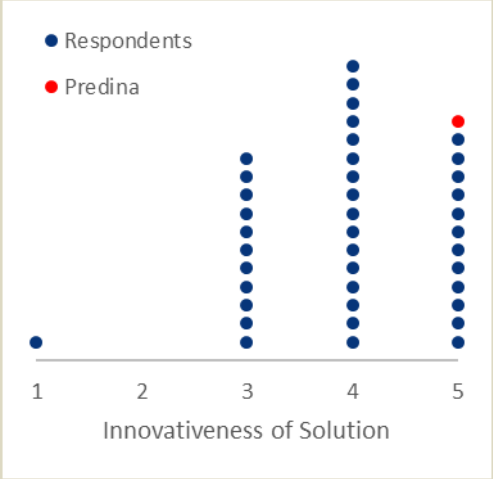
Would you have been able to access the same data without Data Pitch? Yes, had a relationship with Data Provider prior to Data Pitch.



Predina's data provider helped shape the product and turn their algorithm into the commercial solution. Predina note that the credibility this partnership provided for them was also hugely beneficial. Lastly, this project had also provided them a use case for their product.

4.4.4 Acceleration performance

During the accelerator period, Predina also raised additional investment and revenue.[67] Predina also received a sizeable investment from their data provider at the end of the accelerator period.



4.4.5 Since the programme / Outlook

At the start of Data Pitch, Predina defined what successful participation in Data Pitch looked like with the following achievements:

- Change in user behaviour towards risk notification – higher alertness in high-risk situations and less accidents;
- An actual reduction in accidents for UK (and Germany);
- A signed contract for both UK and German markets;
- Real-world testing of Guardian in Germany and other EU countries;
- Scaling of project to full commercial viability and mass-market.

Effort required for scalability:

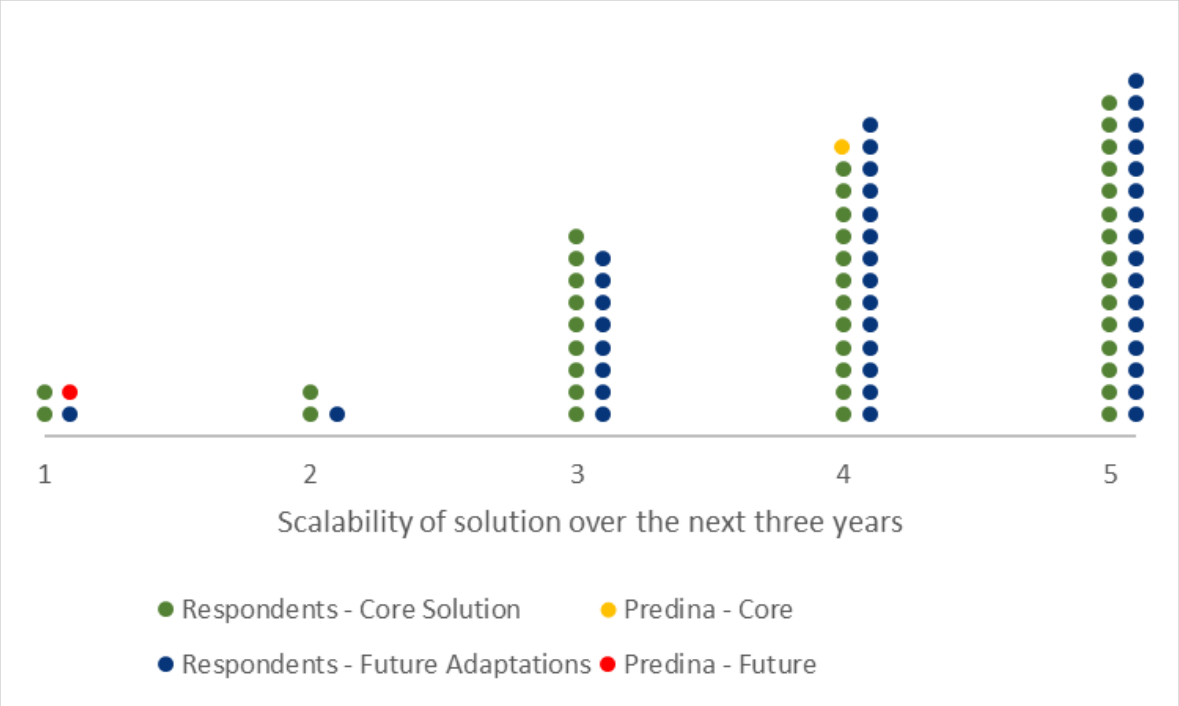
Core solution – Medium

Future adaptations of core solution - Minimum

In the six months after the programme, Predina have managed to show a real-world reduction of 25% in accidents as a result of the implementation of their solution. They were also selected for the Google for Start-ups campus residency programme and enjoyed additional fame in the form of being featured in Forbes 30 under 30 and being selected as Techworld's 'hottest machine learning start-ups' in the UK for the second year running. Predina also increased their employee count by 6 and received additional sales.

In the six months following this period (six to twelve months after the accelerator), Predina have completed version 2 of their technology. They have also secured a pilot with major UK construction companies as well as hiring a new CTO, two additional machine learning engineers and a business development director. They are currently looking for further funding, both grant and strategic investment capital from automotive corporations. Predina is now also looking towards launching in the US.

[67] Precise figures are redacted at Predina's request



Predina's product has been successful both technically (reduction in accidents) and commercially since entering the Data Pitch programme.

4.4.6 Insights

Predina is another example of a start-up with pre-existing relationships with external data providers that used Data Pitch more like a traditional accelerator, with funding and networking being seen as the key benefits of participation.

The opportunities these start-ups had to network with other start-ups was seen as a vital component of the programme for Predina. These opportunities may be a way to encourage further collaboration to provide further impact. When designing the challenges, Data Pitch stated that these challenges should foster discussion and bring together potential solution creators. Future open innovation challenges could examine this area as a potential track for impact generation.

4.5 Radiobotics

Website: https://radiobotics.com/

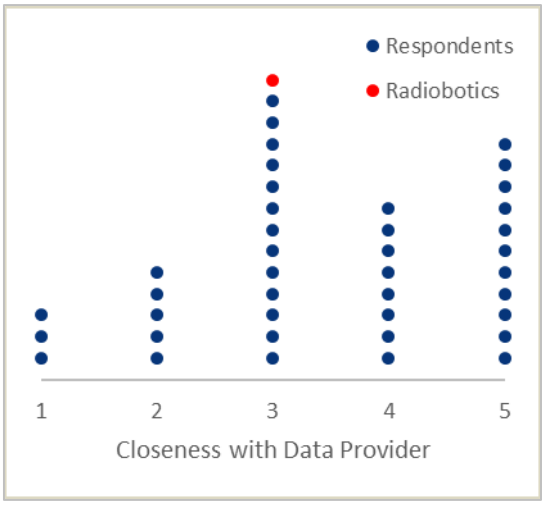
Founded Year: 2017

Number/composition of staff on entry: 2.5 (2 full time, 1 part time)

Challenge: SC1-2017: How can we use data to help people improve their health and wellness and / or make health services more efficient and inclusive

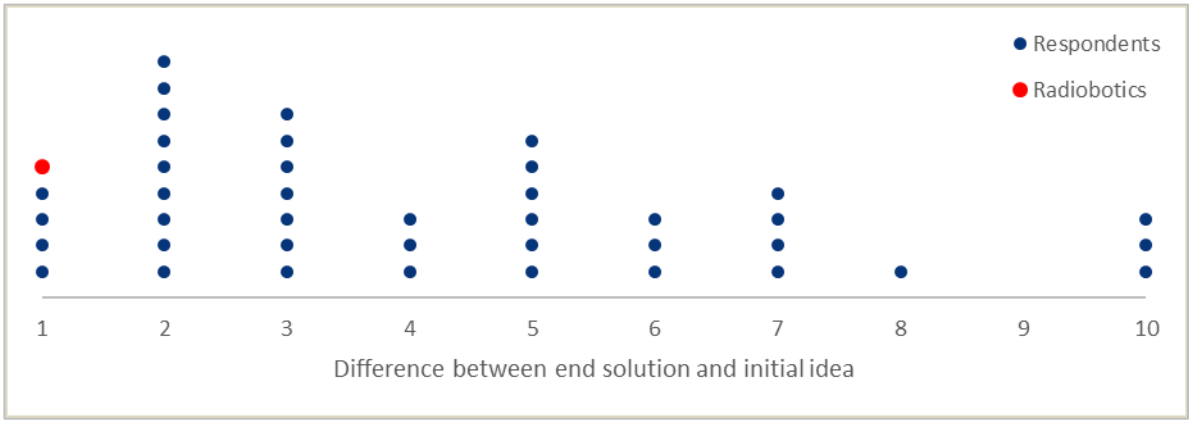
Solution Tagline: Advancing Musculoskeletal Radiology with Machine Learning

4.5.1 Solution



The solution Radiobotics developed as part of Data Pitch involved automating X-ray analysis for hospitals. Specifically, this involved creating machine learning algorithms that could detect osteoarthritis on x-ray images of knees.

Radiobotics aims to sell their solution but will need to receive regulatory clearance due to their solution's status as healthcare technology. Their main goal is to sell their solution to hospitals, with a current focus on European (Denmark, Sweden and UK specifically) and American clients.



4.5.2 Data

As part of their solution, Radiobotics received 20,000 X-ray images. Much of this data was provided by their data provider, although Radiobotics have also received images from both a Danish hospital, and research groups based in the US. Radiobotics received copies of these images and therefore had full control over this data.

Number of Datasets used: 20,000

Update frequency of the data: Occasional/Irregular

Level of control over the data: Full

Size of the data: 5GB, 20,000 observations

This data was used to train their developed algorithm. Radiobotics comment that diversity of images is important to avoid overfitting. Additionally, having multiple data providers ensures a continuous flow of data.

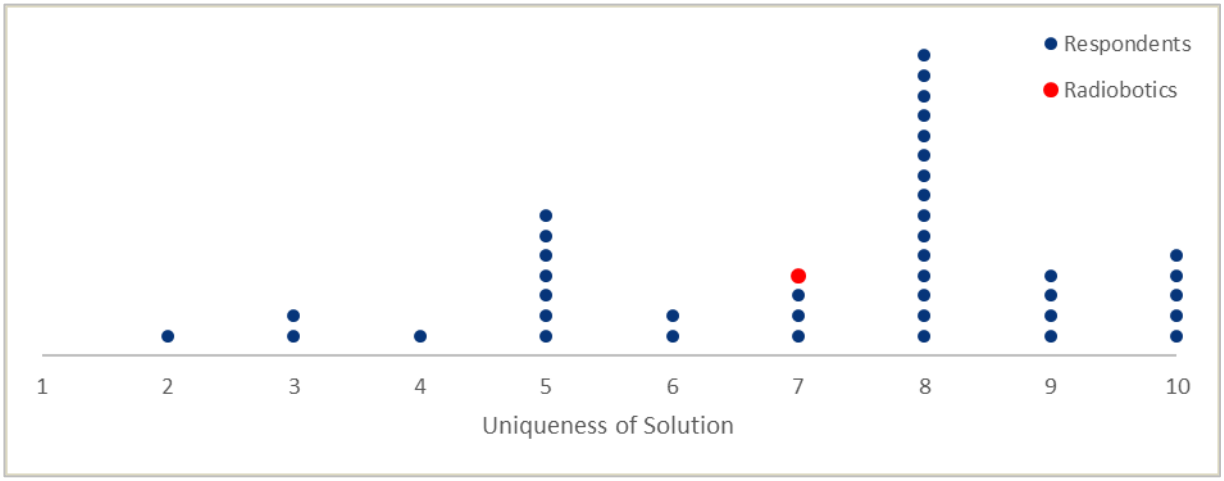
4.5.3 Benefits from the programme

The Data Pitch funding was mostly spent on staffing and salary costs. With this funding, Radiobotics hired 2 new employees.

Resources Data Pitch funding enabled Radiobotics to acquire: Data science/machine learning skills, software development skills, ICT infrastructure/hardware, business management skills, subject matter/domain knowledge, marketing/sales skills, other IT

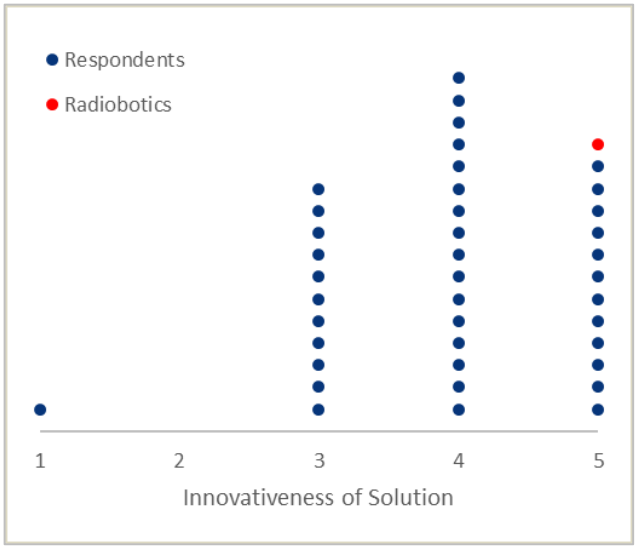
Through Data Pitch, Radiobotics have been able to develop their algorithm to a stage where they could present it to investors. Based on this, they managed to attract additional funding from Eurostars after the Data Pitch programme had ended. Radiobotics appreciate that Data Pitch enabled them to obtain funding without prior traction, as other avenues of funding they had researched required evidence of previous funding, or a product which was to be further developed.

Radiobotics stated that they received support from Data Pitch in the form of events and workshops. The programme has allowed them to both improve their project management skills and accelerate their market and funding analysis as a result.



4.5.4 Acceleration Performance[68]

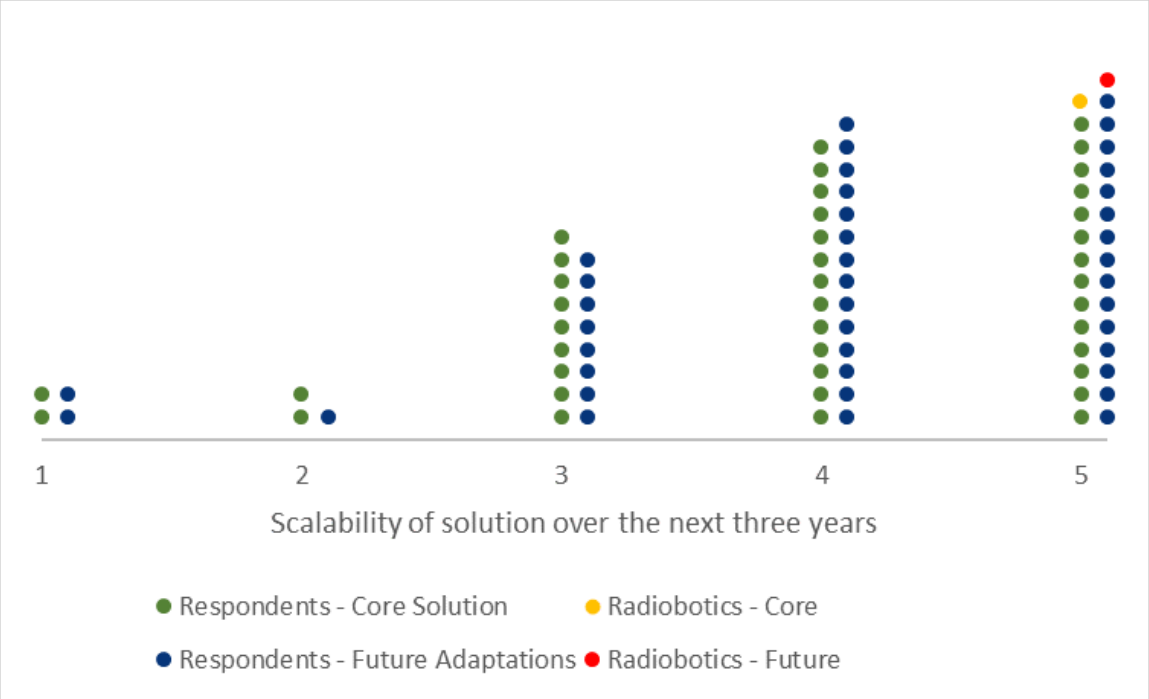
During the accelerator period, Radiobotics raised additional investment. This total amount was split between an investment deal and a grant from the European Institute of Innovation & Technology Health consortium.



4.5.5 Since the programme / Outlook

In the six months after the programme, Radiobotics have increased their staff by an additional two employees. They have also won a substantial amount of investment, including the SME instrument (H2020) phase 1 grant, as well as the aforementioned Eurostars grant. Their Eurostars application had scored the highest out of all previous rounds over the last 10 years.

[68] Approval has not been given by Radiobotics to report precise figures



In the six months following this (between six and twelve months after the end of the accelerator), Radiobotics completed an investment round, obtaining even further investment as a result.

Effort required for scalability:

Core solution – High

Future adaptations of core solution - High

4.5.6 Insights

Radiobotics' solution is aimed at the sensitive and highly regulated healthcare sector, which means that a workable demonstration of the technology is indispensable for gaining traction. The technology developed during the acceleration period with Data Pitch served as a stepping stone for additional funding and contact with investors. Radiobotics specifically highlighted the importance of diverse data sources for their solution to counter the risk of overfitting (a solution that only works for a narrow set of cases).

4.6 Recogn.ai

Website: https://www.recogn.ai/

Founded Year: 2017

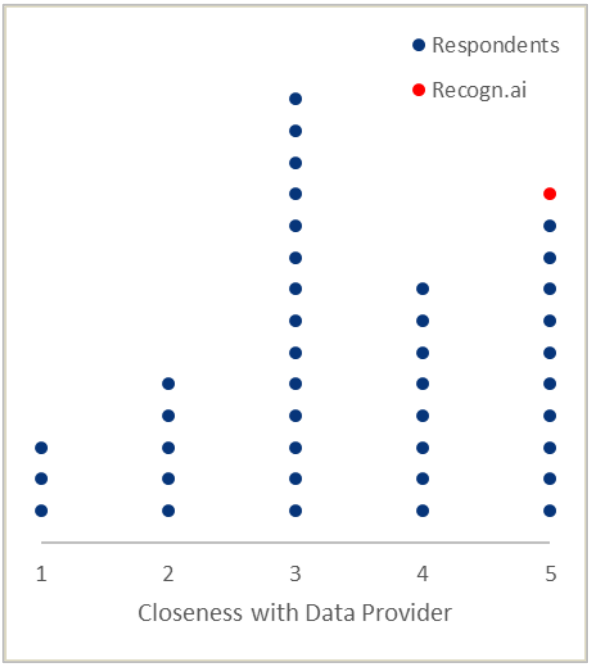
Cohort: 1 – (2017/2018)

Number/composition of staff on entry: 2

Challenge: DPC5-2017: The next generation of customer data management solutions

Solution Tagline: Spin-out from the Technical University of Madrid building AI software for advanced information extraction and integration

4.6.1 Solution

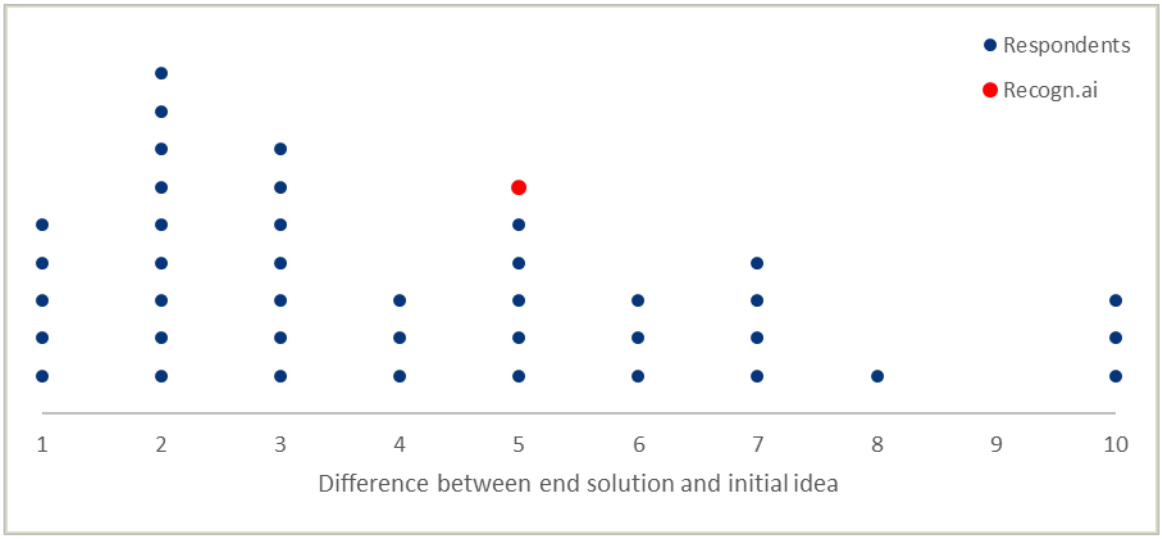


Recogn.ai's solution uses deep learning to match records across different data sources and to create new classifications and categories. The solution can identify customers across transaction datasets and other related customer datasets.

Through the Data Pitch programme, Recogn.ai used customer data received from their data provider Uniserv, in conjunction with open source data sets, to train and test their Machine Learning models. At the end of the programme, their solution could classify records as belonging to a person or business, further allowing classification according to business type.

Recogn.ai's solution also included the development of a UI and ecosystem, referred to as 'Recogn.ai Biome', which allowed for the

creation of model pipelines. Recogn.ai Biome provides an interface which allows users to train a predictive model and register data sources used for training throughout the development process. Additionally, this ecosystem provides comprehensive version control, enabling the user to revert to previous versions, and feedback on model performance through performance metrics such as predictive accuracy.



4.6.2 Data

Number of Datasets used: 10

Update frequency of the data used during the acceleration period: Occasional/Irregular

Update frequency of the data used in the commercial solution: Live/ regular

Level of control over the data: Full

As part of their solution, Recogn.ai used data provided by their data provider, as well as several open datasets. Recogn.ai received 10 GB of data from Uniserv containing over 10 million observations, with information on individuals and businesses in Germany, France and the UK. Uniserv works with datasets provided by their customers. Their customers include businesses such as retailers, banks, financial institutions, insurers and utility providers. Real data relating to businesses were provided to Recogn.ai whereas a synthetic dataset was provided for customer-related data. This customer dataset shared identical statistical properties to the real data set (which could not be shared due to GDPR). Uniserv have stated that this data has commercial value, although a quantifiable figure could not be specified.

The open source datasets used for additional training and testing were sourced from: Wikidata, DBpedia, National Libraries Data, Open Street Maps, Open Addresses.io & Electoral rolls. The team at Recogn.ai had prior experience of working with these and similar datasets. These open data sources totalled an additional 10 million observations and brought the total size of the data used to 20 GB. Recogn.ai had full control over the datasets used in the construction of their solution.

Size of the data received from Data Provider: 10GB, 10,000,000 observations

Size of all other data sets used: 10GB, 10,000,000 observations

4.6.3 Benefits from the programme

Most important resources Data Pitch funding enabled Recogn.ai to acquire: Software development skills; Data science skills; ICT infrastructure, hardware and skills.

Could you have obtained this data (from the Data Provider) without Data Pitch? No

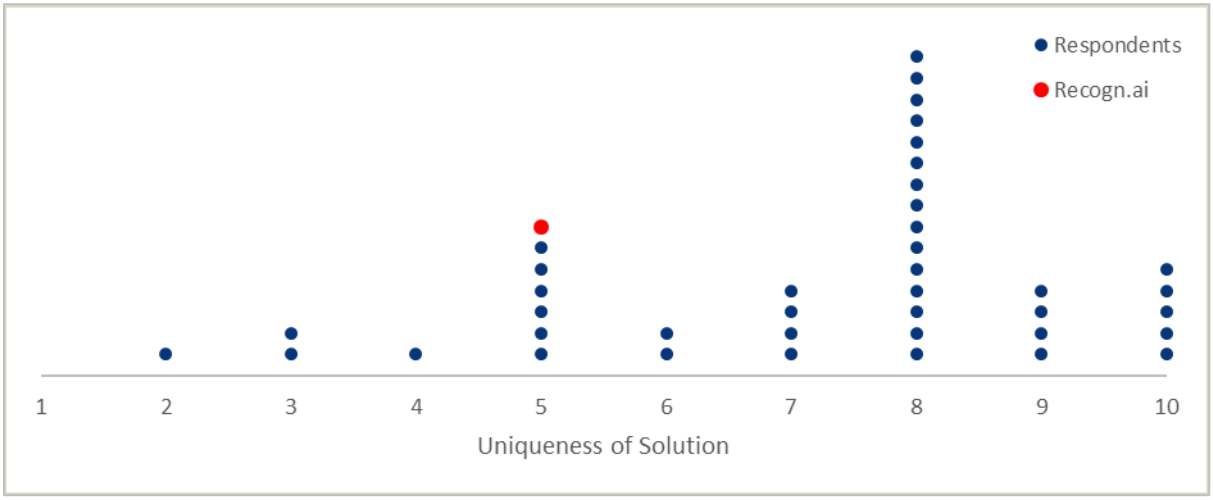
Recogn.ai used the funding they received from Data Pitch:

- to pay the salaries four staff;
- to subcontract a software architect working with the Ontology Engineering Group in Madrid;
- to hire a legal advisor; and,
- to cover travel expenses.

As Recogn.ai had only formed as a company near the start of the project, this funding was vital to allow Recogn.ai to form a team able to develop the product through its initial phases. This initial funding has enabled the company to successfully accelerate through the project, as well as progress the company to its current position.

Support provided by Uniserv provided alongside their data proved to be a major benefit. This included technical information that facilitated working with the data as well as detailed background information on their business objectives and client needs.

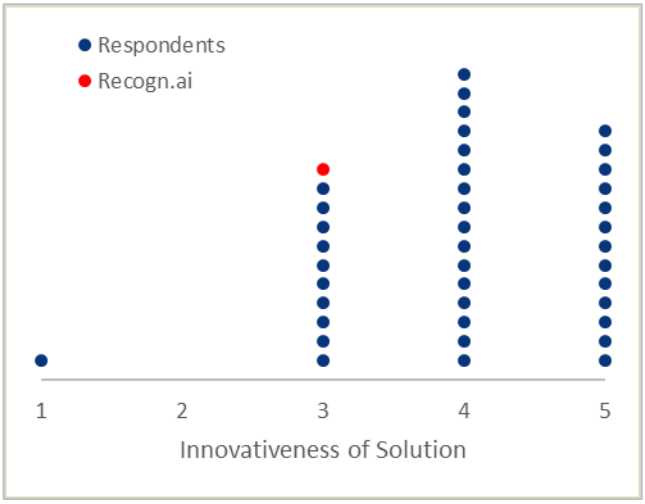
Recogn.ai was also given guidance on GDPR compliance. Recogn.ai highlighted the usefulness of the GDPR workshop Data Pitch had provided for them.



4.6.4 Acceleration Performance[69]

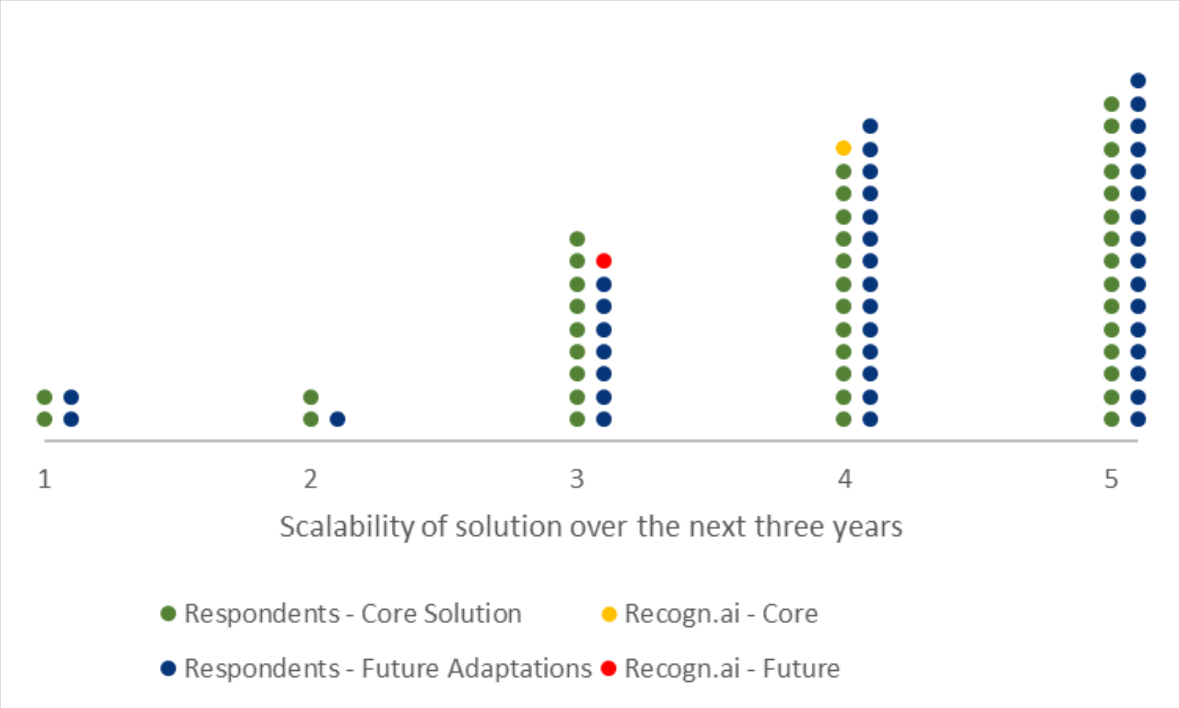
Prior to Data Pitch, Recogn.ai had received investment, but no revenues. During the accelerator period, Recogn.ai both increased their revenues and headcount. Recogn.ai also closed four contracts during the Data Pitch programme. These were:

- A one-year pilot with Correos (Spanish Post Office);
- A 3-month project with the Ministry of Culture and Sport (Spain);
- A 3-month pilot with Evo Banco;
- A six-month project with Zaragoza City Government.



4.6.5 Since the programme / Outlook

In the six months after Data Pitch, Recogn.ai had secured five new clients with whom they are running pilot projects. These include legal firms and insurance companies. Additionally, Recogn.ai had increased their employment by 1, with the addition of a new lead data scientist.



[69] Approval has not been given by recogn.ai to report precise figures

They have also been continuing their work towards attaining a partnership with Uniserv. They were also selected for the Airbus Bizlab Accelerator programme. Their total sales also increased in this six-month period.

Within 12 months after the end of Data Pitch, Recogn.ai had totalled an additional sales, investments and realised efficiencies.

Through Recogn.ai's continued work on their partnership with Uniserv, they have also signed a contract to develop a proof of concept for one of Uniserv's major clients, DZ Bank. Recogn.ai also successfully completed the Airbus Bizlab Acceleration programme and are currently focusing on reaching new clients through a new marketing strategy, as well as continuing with product development and their current projects.

Effort required for scalability:

Core solution – Medium

Future adaptations of core solution - High

4.6.6 Insights

Recogn.ai was a very new company when it joined Data Pitch and another example of a very productive relationship between data provider and start-up. In addition to the matched data provider and the traditional incubator/accelerator roles of Data Pitch, help received in building the team and legal advice were highly appreciated by the start-up.

For a new company, being able to work with a large company can be a rare opportunity. In Recogn.ai's case, their partnership with Uniserv has not only led to further opportunities and deals outside of this partnership but has also provided additional work from a client within Uniserv's network. Therefore, their participation in Data Pitch has had a compounding effect on their business development; Recogn.ai has gone from a pre-revenue start-up, to a company working with banks, government councils and large international companies. When innovation programmes can connect these larger partners to start-ups (whom these larger companies may not be aware of), a beneficial partnership for both sides can be achieved.

5 Conclusions & recommendations

5.1 Conclusions

5.1.1 General success factors of accelerators[70]

In general, the evidence on the success of accelerators is mixed with many accelerators providing different types of outcomes for their participants. However, Bone et al. (2019) show – in a survey of previous participants in accelerator programmes – that start-ups perceived direct funding to be the most useful support received as part of their incubator or accelerator programme. This was followed by access to office space, lab space and technical equipment.

Bone et al. (2019) report stronger evidence for the impact of particular types of support. While most types of support have a significant positive association with at least one of a number of outcome measures (development stage, number of patent applications, R&D expenditure and investment raised, etc.), the strongest evidence for positive impact relates to:

- access to investors;
- access to peers;
- help with team formation;
- direct funding from the programme;
- press or media exposure;
- mentoring from an industry expert;
- help with measuring social impact; or,
- mentoring from a venture capitalist (VC) / angel.

In addition, some support types (e.g. access to peers and coaching/personal development) appear to act directly on improving start-up outcomes. Others seem to work through changing how start-ups approach raising finance, plan strategically, develop and recruit staff, and partner with external organisations. Data Pitch shares the key success factors with other accelerators and is therefore likely to produce similar positive baseline impacts, in addition to benefits that are specific to the unique data sharing aspects of Data Pitch.

5.1.2 Data Pitch success factors

Data Pitch has had substantial positive impacts on participants, both start-ups and data providers. Data Pitch enabled data-driven innovation that would not otherwise have taken place and had a positive impact on the commercial performance of the participating start-ups as measured by increases in employment, revenues and external funding. In addition, Data Pitch succeeded in laying the groundwork for a sustainable open innovation ecosystem in Europe, by providing a platform for data providers to try out open innovation in a low risk setting (facilitating matching and interaction with start-ups and providing expertise and practical support).

Start-ups and data providers identified a number of factors that contributed to the impact of Data Pitch.

[70] See Bone et al. (2019).

Outreach and recruitment of start-ups managed to reach a large, pan-European network of businesses. On the side of data providers, the situation is less clear. Data Pitch may have benefited from a greater number of data providers offering challenges, and a clearer communication between Data Pitch and the data providers regarding required capacity for the programme. A longer and more thorough pre-acceleration phase ("phase 0") may be helpful in the future to ensure that a sufficient number of data providers is available in the project, and to ensure that data providers understand the requirements of the programme. Furthermore, a greater number of providers in future projects may be helpful in limiting the number of start-ups matched to one data provider and ensuring that data providers can dedicate sufficient resources to the solutions developed through open innovation.

Data Pitch succeeded in providing support to a very diverse range of organisations. The programme worked particularly well for start-ups with an existing product or service that could be developed further. Some data providers desired more attention for the data provider. Data providers are typically new to open data innovation and are therefore not always mature enough to make datasets available in ways that allow for immediate work. A longer and more thorough "phase 0" may also help in this respect. Beyond recruitment, this pre-acceleration phase can be used to not only recruit but also onboard and prepare data providers to foster cooperation with start-ups.

Regarding delivery, participants nearly unanimously agreed that administrative burden was minimal and, overall, the delivery of the programme was compared favourably with other accelerators. Data Pitch was especially compared favourably to other publicly funded accelerators.

Both start-ups and data providers recognised that Data Pitch managed to bring together start-ups with skill and data providers interested in open innovation. As such, Data Pitch addressed a gap in the innovation support landscape. Data providers saw clear benefits from the programme in terms of fostering organisational commitments to open innovation, while noting that the success of the open innovation model rests on the individuals involved. The matching inherent to Data Pitch also allowed start-ups to enter into new markets.

The European dimension of Data Pitch was seen as an advantage, especially by start-ups based outside of the traditional start-up hubs. This international dimension allowed start-ups to obtain data which otherwise would have been out of reach. On the flip side, the virtual set-up of the programme allowed for less close contact between start-up and data provider. The data providers especially would have liked closer interaction with their matches start-ups.

Moreover, "anecdotally, incubators and accelerators often serve as 'focal points' for an ecosystem, providing not only a degree of deliberate coordination but also a geographic focus which increases the chance of serendipitous interactions" (Bone et al., 2019, p. 60). The setup of Data Pitch may produce less of these 'serendipitous interactions'. However, the geographic spread of the participants (and the consortium partners) may also in fact widen the scope for possible interactions outside of traditional start-up hubs. It is therefore not clear whether a focus, non-virtual set-up will deliver better results.

5.1.3 Impact of Data Pitch

Impact on start-ups[71]

During the Data Pitch programme, firms increased their sales by a mean of €36,554, received investments of a mean of €82,448, realised efficiencies of €17,168 and increased their employment by an average of 2 employees. On average, start-ups generated €599,432 in sales and €338,862 in investment **per GB of data shared** with them through Data Pitch.

Both ROI and leveraged investment are already substantial during the programme. By the end of the programme, the total Data Pitch resources already attracted 50% equivalent value from other investment opportunities. One year after the end of the programme, ROI already exceeded 100% and leveraged investment was already approaching 300%.

Table 14 Return on Investment and leveraged investment; realised and projected figures

	During the programme	6 months following the programme	12 months following the programme	Projected annual figure by 2022
Return on Investment	23%	91%	103%	459%
Leveraged investment	50%	82%	278%	N/A

Note: data on 6 and 12 months after the programme are based on data from Cohort 1 only. The figures account for this by adjusting the investment provided through Data Pitch based on the number of start-ups for which data is available.

Source: Bi-weekly monitoring updates, 6 months progress update, 12 months progress update, revenue growth projection

The majority of start-ups would not have been able to access the same or similar data without Data Pitch., in particular start-ups working in the financial and medical sectors. Access to data does seem to influence the ability of start-ups to attract additional funding. Start-ups that could access data outside of Data Pitch attracted, on average, €141,000 more in additional funding than start-ups that would not be able to access data. Start-ups typically had full control over the data when building the solution. Similarly, data was typically stored under the control of the start-ups.

The majority of start-ups used machine learning in their solution. The outcomes tracked during the programme provide some evidence that using machine learning helps attract investment. There is an impressive difference between start-ups that use machine learning and those that do not. Start-ups that did use machine learning attracted, on average, €108,000 more in investment than that did not use machine learning.

Regarding growth opportunities, start-ups in the sector challenges had higher growth expectations for their product. This (perceived) ability to scale a solution has a distinct impact on the ability to attract funding; with easier perceived ability to scale being associated with increases in additional funding received.

Successful applicants received, on average, more external funding than unsuccessful applicants (not including the €100,000 received through Data Pitch). Some applicants, both successful and unsuccessful, were able to attract substantial investments upwards of €500,000.

[71] Note that impacts might not have been fully disclosed by the participating start-ups (or materialised) by the end of the programme. So that the impact may be understated somewhat.

Projecting performance of funded start-ups into the future, a forecast model predicts that average revenue for start-ups will grow from €147,723 in 2019 to €833,555 in 2022. Combined with the success/failure rate of businesses, this implies a growth of total annual revenue to €35,784,385 from €6,896,000. This is equivalent to a growth of 73% per year up until 2022.

Impact on data providers

The impact on data providers has been reported mostly in terms of promoting the open innovation approach within the provider organisations. All interviewees saw themselves participating in open innovation in the future, several felt they had absorbed sufficient knowledge to do so without external help. Several remarked especially that they were now further along in their journey to make their data more accessible and understandable for third parties. Some data providers mentioned that more contact with other data providers would have been useful. Creating a network of data providers that continues after the acceleration phase might be a useful extension of a programme like Data Pitch.

Although Data Providers went into Data Pitch with a business challenge, they saw Data Pitch as a learning opportunity regarding Open Innovation and data sharing. As such, measuring quantitative impacts were not prioritised and generally not available. However, one data provider estimated that the solution developed with their data in Data Pitch reduced the cost of a particular business process by 35%.

5.1.4 Longer term impacts

Data Pitch is likely to produce a number of impacts in the longer term. These include the future successes of the participating start-ups: Bone et al. (2019) state that "accelerators and incubators affect start-ups in numerous ways. Outcomes of such interactions manifest themselves sometimes immediately and sometimes only a considerable time after a start-up has graduated from a programme". One effect that is pivotal for long-term success but likely too early to evaluate is the effect on firm survival. It has been found that start-ups that graduated from accelerator programmes have approximately 23% higher survival rate than other new businesses (Regmi et al., 2015).

On top of this, there is the impact on data providers and the wider open innovation ecosystem that Data Pitch is helping to initialise:

- The main benefit of Data Pitch for data providers was the learning experience from the programme. It is therefore likely that more tangible impacts as a result of participating in data-driven open innovation will only emerge over time.
- Ecosystem effects in relation to accelerators, and specifically in relation to open innovation, require evaluation over the longer term. This would include both the effect of Data Pitch on the wider start-up support ecosystem in Europe and the synergies between Data Pitch and other forms of support available to start-ups.

5.2 Recommendations

- Sector challenges and provider challenges seem to have worked differently. For example, start-ups in the data providers challenges and the sector challenges differed in the methodologies used (sector challenge start-ups were more likely to use machine learning), the type of data used (sector challenge start-ups used video data, whereas data provider challenge start-ups did not) and the way data was stored (data provider challenge start-ups mostly files, whereas sector challenge start-ups mostly used relational databases) as

shown in section 3.6. A more exclusive focus on matched challenges is likely to produce greater benefits and provides a better testbed for the hypothesis that reducing frictions inhibiting data sharing facilitates can unlock data-driven open innovation. The experience for participants in the sector challenges more closely resembled that of a standard accelerator, where the start-up is able to source data from external parties independently.

- More resources to prepare and connect data providers may maximise the programme impact. This could involve a "Phase 0" with a selection process and an 'acceleration' period focused on data providers to prepare them for working with start-ups on their challenges. The long-term benefits for data-driven open innovation could be cemented by the creation of a network of data providers that persists after the end of the programme.
- Despite the comprehensive support provided to the start-ups by Data Pitch, the more mature start-ups seem to have performed better. A clearer focus on start-ups with 'acceleration-stage' maturity (proven ability to deliver an MVP) may enhance the overall impact. Partly this could reflect the fact that more experienced start-ups are better able to select appropriate challenges both based on their technical viability, but also strategically, in terms of the applicability of an existing solution to a new market or industry. In this regard, a stricter separation between 'incubation-stage' start-ups, who may not be able to produce a working prototype by the end of the programme, and the more mature start-ups may be contemplated, so as to provide each type of start-up with the optimal support package, with more strategic advice being more appropriate for the more mature start-ups.
- There is some evidence that the impact of Data Pitch was stronger in sectors with higher barriers to data sharing (such as healthcare and finance, which data subjects see as particularly sensitive). For example, as discussed in section 3.7.2, start-ups in healthcare and finance were more likely to note that, without Data Pitch, they would not be able to access data. An ex-ante focus on such sectors may increase benefits. The selection of sectors should take into account their specific barriers to, and enablers of data sharing. For example, data-driven open innovation in the finance sector is supported by strong regulatory action (Open Banking, PSD2).
- The programme design and setup should facilitate robust evaluation of the programme itself. This includes having a clear evaluation strategy with well-defined success metrics for the programme, and comprehensive baseline data collection. Making long-term data sharing obligatory for participants may be considered. Which data is shared long-term, which may be a couple of years, depends on the success metric targeted by the programme, but may include revenues and employment figures of participants.
- More broadly, investment in further pilot projects is needed to develop the open innovation model and to find out what works. Parameters such as the selection of start-ups and type of support given should be comparatively analysed. An opportunity exists with using other European incubators, such as the European Data Incubator. Other European programmes may be used to experiment with, for instance, on-boarding processes in similar, but not identical, circumstances.

References

Bone, J. Allen, O. & Haley, C. (2017). *Business Incubators and Accelerators: The National Picture.* BEIS Research Paper Number 7. Retrieved from: https://www.gov.uk/government/publications/business-incubators-and-accelerators-the-national-picture

Bone, J., Gonzalez-Uribe, J., Haley, C. & Lahr, H. (2019). *The Impact of Business Accelerators and Incubators in the UK.* BEIS Research Paper Number 2019/009. Retrieved from: https://www.gov.uk/government/publications/the-impact-of-business-accelerators-and-incubators-in-the-uk

Chesbrough, H. & Bogers, M. (2014). *Explicating Open Innovation: Clarifying an Emerging Paradigm for Understanding Innovation.* In H. Chesbrough, W. Vanhaverbeke & J. West (Eds.), *New Frontiers in Open Innovation.* Oxford: Oxford University Press, pp. 3-28.

Churchill, N.C. & Lewis, V.L. (1983). *The Five Stages of Small Business Growth.* Harvard Business Review. Retrieved from: https://hbr.org/1983/05/the-five-stages-of-small-business-growth

Clarysse, B., Wright, M. & Van Hove, J. (2015). *A Look Inside Accelerators: Building Businesses.* Retrieved from: https://www.nesta.org.uk/report/a-look-inside-accelerators/

Coadec & The Entrepreneurs Network (2019). *The Start-up Manifesto.* Retrieved from: https://static1.squarespace.com/static/58ed40453a04116f46e8d99b/t/5de921a956a00a7f2764220f/1575559602818/The+Start-up+Manifesto.pdf

Ctrl-Shift (2018). *Data Mobility: The personal data portability growth opportunity for the UK economy.* Report for the Department for Digital, Culture, Media & Sport. Retrieved from: https://www.ctrl-shift.co.uk/reports/DCMS_Ctrl-Shift_Data_mobility_report_full.pdf

IDC (2017). *Impact Assessment of ODINE Programme.* Retrieved from: https://opendataincubator.eu/wp-content/uploads/2016/01/ODINE-Final-report-by-IDC.pdf

HM Treasury (2018). *The Green Book: Central Government Guidance on Appraisal and Evaluation.* Retrieved from: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

Miller, P. & Bound, K. (2011). *The Start-up Factories: The rise of accelerator programmes to support new technology ventures.* Nesta Discussion Paper. Retrieved from: https://media.nesta.org.uk/documents/the_start-up_factories_0.pdf

OECD (2005). *Glossary of Statistical Terms: Innovation.* Retrieved from: https://stats.oecd.org/glossary/detail.asp?ID=6865

Puri, M., Zarutskie, R. (2012). On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firm. The Journal of Finance, 67(6), 2247-2293; Table VI

Regmi, K., Ahmed, S.A. & Quin, M. (2015). *Data Drive Analysis of Start-up Accelerators.* Universal Journal of Industrial and Business Management, 3(2), pp. 54-57. DOI: 10.13189/ujibm.2015.030203

Index of tables & figures

Tables

Figures

Annex 1 List of Data Pitch challenges

Table 15 Call 1 challenges (July 2017)

Challenge Identifier	Sector	Challenge	Data provider
DPC1 - 2017	RETAIL	Future-proof retail supply chains	Sonae Center Servicos II, S.A.
DPC2 - 2017	SPORTS & RECREATION	How can we use data to improve visibility and access to physical activities?	Imin Limited
DPC3 - 2017	DATA ANALYTICS	Empowering sales and marketing decisions through company knowledge graphs	SpazioDati
DPC4 - 2017	TRANSPORT	Changing public transport for the better	Deutsche Bahn AG
DPC5 - 2017	DATA MANAGEMENT	The next generation of customer data management solutions	UniServe GmbH
SC1 - 2017	HEALTH & WELLNESS	How can we use data to help people improve their health and wellness and/or make health services more efficient and inclusive?	N/A
SC2 – 2017	EMPOWERING USERS ONLINE	How can we use data to make the Web more trustworthy and improve personal safety and security online?	N/A
SC3 – 2017	LIFELONG LEARNING	How can we use data to ensure that we have and can further develop the skills we need in the future?	N/A
SC4 – 2017	LIVING	How can we use data to improve living standards and lifestyle, and create new accommodation options in Europe?	N/A
SC5 – 2017	SMART MANUFACTURING	How can we use data to make manufacturing, logistics and maintenance processes more efficient and able to support new models of use and repair?	N/A
SC6 - 2017	TOURISM	Transforming tourism: aggregated travel services and intelligent personal assistants	N/A
OIC - 2017	OPEN INNOVATION	Harnessing the full power of data-driven innovation	N/A

Note: = Provider challenges

Source: Data Pitch deliverable D4.2.: Summary of round 1

Table 16 Call 2 challenges (July 2018)

Challenge identifier	Sector	Challenge	Data provider
DPC1-2018	Personalised entertainment	Developing the next generation of multidimensional recommendations	Altice Labs SA
DPC2-2018	Text mining and analytics	Automated answering of subjective questions on environmental and social governance	Bloomberg
DPC3-2018	Smart manufacturing	Developing applications across manufacturing, logistics and supply chain	Greiner International Packaging
DPC4-2018	Sustainable food supply chain	Creating farm-to-market linkages	GROW

Challenge identifier	Sector	Challenge	Data provider
DPC5-2018	Customer needs predictions	Creating adaptive ways to anticipate customer requirements	Konica Minolta
DPC6-2018	Healthcare	Creating outcome-based healthcare offerings	Jose de Mello Saude
DPC7-2018	Multimodal transport	Seamless travel services across Europe	MASAI
DPC8-2018	Weather and climate change	Creating social and economic value by reducing the impact of climate change	MET Office
SC1-2018	Pharmaceuticals	Developing innovative approaches and processes across the pharmaceutical industry	N/A
SC2-2018	Automotive	Maximising the positive impact of autonomous connected, electrified and shared vehicles	N/A
SC3-2018	Energy	Increasing efficient energy creation and use	N/A
SC4-2018	Finance	Overcoming the data challenges in the financial sector	N/A
SC5-2018	Telecoms	Supporting 5G readiness and deliver tomorrow's telecoms industry	N/A
SC6-2018	Privacy and consent control	Creating products and services to ensure individual privacy and control	N/A
SC7-2018	Smart transport	Innovative solutions to improve mobility and reduce traffic congestion	N/A
OC1-2018	Open challenge	Harnessing the full power of data-driven innovation	N/A

Note: = Provider challenges

Source: Data Pitch deliverable D4.3.: Summary of round 2

Annex 2 Impact forecast methodology

A2.1 Employment growth

A2.1.1 Model and assumptions

Employment growth was estimated using a simulation-based approach with random model inputs. That is, the model defined a calculation of employment growth based on a set of input assumptions. The calculation drew these inputs at random from a pre-defined distribution. This exercise was repeated in total 10,000 times, drawing different random inputs at every repetition, and the average over these 10,000 simulations provides the final estimate of growth.

The same model was used to estimate a forecasted growth model and a counterfactual model. The counterfactual model provides an estimate of employment growth had Data Pitch not existed. The forecast and counterfactual models differed only in the input assumptions used. The core functionality of the model was the same.

The model combined data from both cohorts into one single cohort and started the projection at the start of this hypothetical, combined cohort. For the sake of reporting, the start of the combined cohort is equated with the start of cohort 2. Therefore, baseline figures refer to the year 2019, and the subsequent years refer to respectively 2020, 2021 and 2022.

A2.1.2 Baseline assumptions

The model firstly required a baseline from which to evaluate growth in employment. The assumptions for this baseline were the same for the forecast and the counterfactual model, and were:

- there are 47 SMEs at the start of the Data Pitch programme;
- they have on average 7[72] employees at the start; and,
- they have 310 employees in total at the start.

The average and total number of employees for the 47 SMEs was derived from the stated employment of successful applicants during the application stage.

A2.1.3 Core functionality of the model

The model calculated the following three outcomes for the years 2020, 2021 and 2022:

- average employment per start-up;
- number of start-ups still in business; and,
- total employment for all start-ups still in business.

This was calculated as follows:

[72] Note that the unrounded average is 6.58, which explains why total employment does not equal $47 * 7$.

For 2020 (the first year after the programme):

- average employment was calculated as $7 * (1 + baseline\ employment\ growth\ rate)$ where the baseline growth rate was a randomly drawn input (see below);
- the number of start-ups in existence as calculated as $47 * (1 - startup\ death\ rate)$ where the start-up death rate was also randomly drawn; and,
- total employment was calculated as $average\ employment_{2020} * existing\ startups_{2020}$.

For 2021:

- average employment was calculated as $average\ employment_{2020} * (1 + 2021\ employment\ growth\ rate)$ where the 2021 growth rate was a function of the baseline growth and a scaling factor (see below);
- the number of start-ups was calculated as $existing\ startups_{2020} * (1 - startup\ death\ rate)$; and,
- total employment was calculated as $average\ employment_{2021} * existing\ startups_{2021}$.

Finally, for 2022:

- average employment was calculated as $average\ employment_{2021} * (1 + 2022\ employment\ growth\ rate)$ where the growth rate was defined alongside the 2021 growth rate;
- the number of start-ups was calculated as $existing\ startups_{2021} * (1 - startup\ death\ rate)$; and,
- total employment is calculated as $average\ employment_{2022} * existing\ startups_{2022}$.

Note that number of employee and businesses cannot be fractional. Both were rounded to the nearest integer during the calculations.

A2.1.4 Baseline employment growth rate

The inputs used for the baseline employment growth rate in the forecast model were derived from the bi-weekly monitoring updates and 6 month progress update. Where both were available, the data from the updates were summed to derive 12-month progress since the start of the programme. Where the 6 months update was not available (cohort 2), the bi-weekly updates were multiplied by 2 to derive an annualised figure.

The reported changes in employment were mapped to six categories[73]. These categories themselves were mapped into implied baseline growth rates. The probability distribution with which these baseline growths were applied in the simulations, was derived from the proportion of start-ups in each of the six categories. The assumption on baseline employment growth in the forecast model was therefore as follows:

[73] Change in employment was categorised into a discrete number of categories to ensure that the same model can be used for both the forecast and counterfactual scenario.

Table 17 Baseline employment growth rates in the forecast model

Category	Implied baseline growth[a]	Probability
More than 15 employees more in first year	214%	2%
Between 11 and 15 employees more in first year	186%	2%
Between 6 and 10 employees more in first year	114%	32%
Between 1 and 5 employees more in first year	43%	40%
No change in employment in first year	0%	19%
Between 1 and 5 employees fewer in first year	-43%	4%

[a] Implied growth rates have been calculated as follows:

More than 15: $(15/7) * 100\% \approx 214\%$

Between 11 and 15: $(13/7) * 100\% \approx 186\%$

Between 6 and 10: $(8/7) * 100\% \approx 114\%$

Between 1 and 5: $(3/7) * 100\% \approx 43\%$

No change: 0%

Between 1 and 5 decrease: $(-3/7) * 100\% \approx -43\%$

Except for "More than 15" and "No change", the growth rates are based on the mid-point of each category. Note that the average employment was 7 at the start of the programme.

Source: Bi-weekly monitoring updates and 6 months progress update

For the acounterfactual scenario, the inputs in the model were based on the 'unsuccessful applicants survey'. This survey gauged change in employment in 5 categories which were similarly mapped to implied baseline growth rates. The probability distribution with which these baseline growths were applied in the simulations, was derived from the proportion of start-ups in each of the five categories. The assumption for the counterfactual model was therefore as follows:

Table 18 Baseline employment growth rates in the counterfactual model

Category	Implied baseline growth[a]	Probability
More than 10 employees more in first year	143%	22%
Between 6 and 10 employees more in first year	114%	0%
Between 1 and 5 employees more in first year	43%	56%
No change in employment in first year	0%	11%
Between 1 and 5 employees fewer in first year	-43%	11%

[a] Implied growth rates have been calculated as follows:

More than 10: $(10/7) * 100\% \approx 143\%$

Between 6 and 10: $(8/7) * 100\% \approx 114\%$

Between 1 and 5: $(3/7) * 100\% \approx 43\%$

No change: 0%

Between 1 and 5 decrease: $(-3/7) * 100\% \approx -43\%$

Except for "More than 15" and "No change", the growth rates are based on the mid-point of each category. Note that the average employment was 7 at the start of the programme.

Source: Unsuccessful applicants survey

A2.1.5 Growth rates for 2021 and 2022

The employment growth rates in 2021 and 2022 were defined as a function of the baseline growth rate and the assessment of scalability of solutions reported in the 'successful applicants survey'. More precisely, successful applicants were asked to rate the ease with which their core solution could be scaled in the three years following the programme from 1 (limited growth) to 5 (significant growth). It is to be expected that start-ups that foresee significant growth will grow faster than start-ups that only foresee limited growth.

The employment growth rates in 2021 and 2022 were thus defined as:

- 2021: $baseline\ growth\ rate * scaling\ factor_{2021}$
- 2022: $baseline\ growth\ rate * scaling\ factor_{2022}$

The scaling factors for 2021 and 2022 were defined as:

- $scaling\ factor_{2021} = \frac{1}{2*(6-scalability)}$
- $scaling\ factor_{2022} = \frac{1}{3*(6-scalability)}$

where $scalability$ is a number between 1 and 5, where 1 indicates limited growth potential and 5 indicates significant growth potential. The probability distribution over $scalability$ was derived from responses in the successful applicants survey. The same assumption was used in the forecast and counterfactual model.

The assumption on the scaling factors (and by extension the 2021 and 2022 growth rates) were thus:

Table 19 Scaling factors for 2021 and 2022 growth rates

Scalability	Scaling factor: Year 2	Scaling factor: Year 3	Probability
1 (limited growth)	0.10	0.07	5%
2	0.13	0.08	5%
3	0.17	0.11	22%
4	0.25	0.17	32%
5 (significant growth)	0.5	0.33	37%

Source: Successful applicants survey

Note that the scaling factors always decreased the (absolute value of) baseline growth. This was implemented because the baseline growth rates could be substantial and in some cases were even above 100%. Such high growths rates are not sustainable in even the relatively short run of three years. As such, growth rates were diminished over time to generate a more realistic assumption. It should be noted that in reality employment growth rates may show more erratic patterns depending on specific company performance.

Furthermore, the decline in growth was larger between 2020 and 2021 than between 2021 and 2022. This accounted for the fact that lower growth rates have less scope for further decline. In effect, if employment growth is 50%, it can reasonably be expected to decrease more than if the growth rate is 25%. By construction, employment growth rates were lower in 2021 than in 2020.

A2.1.6 Start-up death rate

The last input into the model was the start-up death rate. Death rates were collected from a variety of sources. Each death rate was used with equal probability in the simulation and differed between the forecast and counterfactual model.

For the forecast model, the following assumption was used:

Table 20 Start-up death rates in the forecast model

	Death rate	Probability
Death rate of all companies; Eurostat	8%	25%

	Death rate	Probability
Death rate of companies between 5-9 employees; Eurostat[a]	2.48%	25%
Death rate of companies with 10 or more employees; Eurostat[b]	1.29%	25%
Current state among successful applicants	0%	25%

[a] The average number of employees at the start of the programme was 7

[b] In most cases, the size of SME will exceed by 2020.

Source: Eurostat, 2017 (bd_9bd_sz_cl_r2)

The average death rate used in the model was 2.9%. This can be compared with the death rate of VC-funded companies as reported by Puri & Zarutskie (2012). They report a failure rate of 4.9% one year after a VC-funded firm is matched with a venture capitalist. The assumed death rate for the simulation may therefore be somewhat on the low side. However, it should be noted that the profile of VC-funded firms does not necessarily reflect the profile of SMEs funded through Data Pitch.

For the counterfactual model, the following assumption was used:

Table 21 Start-up death rate in the counterfactual model

	Death rate	Probability
Death rate as established by an internet sweep of unsuccessful applicants	13%	25%
Death rate as reported in the unsuccessful applicants survey	11%	25%
Death rate of all companies; Eurostat	8%	25%
Death rate of companies between 5-9 employees; Eurostat[a]	2.48%	25%

[a] The average number of employees at the start of the programme was 7

Source: Eurostat, 2017 (bd_9bd_sz_cl_r2); Unsuccessful applicants survey

The average death rate used was 8.62%. Compared with Puri & Zarutski (2012), this is again on the lower side. The authors report a failure rate of 9.5% in the first year after matching with a venture capitalist.

A2.2 Revenue growth

A2.2.1 Changes in assumptions relative to employment growth model

Most of the assumptions used in the employment growth simulation were also applied to the revenue growth model. Furthermore, the core functionality was not changed. The following assumption were changed.

In the baseline assumption, the average revenue at the start of the Data Pitch programme was assumed to be €146,723.40, which was the average revenue reported by successful applicants at the application stage.

The assumption for baseline revenue growth in the forecast model was again based on data from monitoring updates and the 6 months progress updates. Where available, revenues (including efficiency gains; see notes to Figure 35) were summed across both updates. For cohort 2, revenues were multiplied by 2 to derive annualised figures. Revenues were categorised into five groups. These groups were mapped into implied baseline growth rates. The assumption for baseline growth in the forecast model was as follows:

Table 22 Baseline revenue growth in the forecast model

Category	Implied baseline growth[a]	Probability
Revenue increase by more than €1 million	682%	4%
Revenue increase by between €500,001 and €1 million	511%	2%
Revenue increase by between €100,001 and €500,000	170%	30%
Revenue increase by between €1 and €100,000	34%	28%
No change	0%	36%

[a] Implied growth rates have been calculated as follows:

More than €1 million: $(1{,}000{,}000/146{,}723.40) * 100\% \approx 682\%$

Between €500,001 and €1 million: $(750{,}000/146{,}723.40) * 100\% \approx 511\%$

Between €100,001 and €500,000: $(250{,}000/146{,}723.40) * 100\% \approx 170\%$

Between €1 and €100,000: $(50{,}000/146{,}723.40) * 100\% \approx 34\%$

No change: 0%

Except for "More than €1 million" and "No change", the growth rates are based on the mid-point of each category

Source: Bi-weekly monitoring updates

For the counterfactual model, the inputs were derived from the unsuccessful applicants survey using the same categories as the forecast model. The assumption in the counterfactual model was as follows:

Table 23 Baseline revenue growth in the forecast model

Category	Implied baseline growth[a]	Probability
Revenue increase by more than €1 million	682%	11%
Revenue increase by between €500,001 and €1 million	511%	11%
Revenue increase by between €100,001 and €500,000	170%	22%
Revenue increase by between €1 and €100,000	34%	22%
No change	0%	33%

[a] Implied growth rates have been calculated as follows:

More than €1 million: $(1{,}000{,}000/146{,}723.40) * 100\% \approx 682\%$

Between €500,001 and €1 million: $(750{,}000/146{,}723.40) * 100\% \approx 511\%$

Between €100,001 and €500,000: $(250{,}000/146{,}723.40) * 100\% \approx 170\%$

Between €1 and €100,000: $(50{,}000/146{,}723.40) * 100\% \approx 34\%$

No change: 0%

Except for "More than €1 million" and "No change", the growth rates are based on the mid-point of each category

Source: Unsuccessful applicants survey

Annex 3 Data collection

A3.1 Summary

LE conducted interviews with Data Pitch participants (data providers and start-ups) and ran online surveys of participating start-ups and unsuccessful applicants. A total of 23 interviews were conducted. The online surveys elicited a combined 50 complete responses. Formal consultations were also held with Data Pitch staff from the ODI and the University of Southampton.

Table 24 Primary data collection programme (October – November 2019)

Interviews (face-to-face, phone)	Online survey responses
Start-ups: 17 Cohort 1: 5 Cohort 2: 12	Data Pitch participants: 41
Data providers: 6 Cohort 1: 1 Cohort 2: 5	Unsuccessful applicants: 9

The following sections reproduce the interview guides and survey instruments used in the data collection programme.

A3.2 Data Pitch evaluation – Start-up topic guide

Background

- Background to the SME/solution (did the idea precede participation?)
- How did they become aware of Data Pitch
- Motivation for participating in Data Pitch (and provider vs. sector challenge)
- Potential alternatives to Data Pitch (other forms/sources of funding)

Details of the data provided

- Data type, characteristics (volume, periodicity, granularity, source(s), personal vs non-personal, etc.)
- Benefits of accessing this data through Data Pitch
  - Didn't know this data existed
  - Couldn't locate a provider of this data
  - Too expensive
  - Technical barriers
  - Legal/regulatory barriers

Details of the solution

- Description
- Market positioning, user base
- Use of technology (AI)
- Importance of data in the solution
  - Quantity
  - Diversity
  - Innovation
- Role of open data/open innovation in the solution

Use of the funding received

- What are the funds used for (skills, hardware, etc.)

Other support received from the data provider(s) & Data Pitch

- Any other resources provided to the SME as part of Data Pitch by the data provider
- Any other resources provided to the SME as part of Data Pitch by Data Pitch (e.g. legal documentation)

Benefits

- Direct benefits to the SME
- Benefits to other parties
- Plans for future use of the solution

- Future participation in open innovation

Reflections on the programme

- Application process
- Selection/matching process
- Monitoring arrangements
- Outcomes vs expectations,
- Lessons learned
- Comments & suggestions

Other

- Did you receive/fill in the online survey? Any issues with the online survey?

A3.3 Data Pitch evaluation – Data provider topic guide

Background

- Background to the challenge/innovation sought
- Motivation for participating in Data Pitch
- Potential alternatives to Data Pitch (in-house development, ordinary procurement)

Details of the data provided

- Data type, characteristics (volume, periodicity, granularity, source(s), personal vs non-personal, etc.)
- Value of data (any internal valuation of the data provided)

Other support provided

- Any other resources provided to the SME as part of Data Pitch

Benefits for the data provider

- Direct benefits to the provider
- Benefits to other parties (SMEs, third parties)
- Plans for future use of the solution
- Future participation in open innovation

Reflections on the programme

- Interactions with the Data Pitch consortium
- Selection/matching process
- Outcomes vs expectations,
- Performance of the SME
- Lessons learned

A3.4 Data Pitch evaluation – successful participants survey

London Economics [https://londoneconomics.co.uk] has been commissioned by the Data Pitch consortium to carry out an independent assessment of the Data Pitch programme. This survey collects information on your interaction with Data Pitch and the characteristics and features of your solution. No personal information is being collected in this survey. The survey will take approximately 15 minutes to complete.

If you have any questions, please contact Moritz Godel, T +44 (0)20 3701 7708, mgodel@londoneconomics.co.uk

Q1. Name of your business: *

Q2. How closely did you interact with Data Providers other than your partnered Data Provider? (1 signifies 'No interaction' and 5 signifies 'Very close interaction')

Q3. How many datasets do you use in your solution? Please provide a total number of open, closed and self-generated datasets. (Dataset refers to sets of data that share the same features/characteristics and which your business either receives from a data provider or collects itself). *

Q4. Without Data Pitch, would you have been able to access the same data (or equivalent data that would enable you to implement the same solution)? *

- [ ] Yes
- [ ] No

Q5. What would prevent you from accessing the same data (or equivalent data that would enable you to implement the same solution)? *

- [ ] I didn't know this data existed
- [ ] I couldn't locate a provider of this data
- [ ] Too expensive

- [ ] Technical barriers
- [ ] Legal/regulatory barriers
- [ ] Other (please specify):

Q6. Which characteristics of the data used in Data Pitch are the most important for your solution?*

- [ ] Volume of data (enabling more precise predictions, greater coverage, etc.)
- [ ] Richness/granularity of data (enabling higher quality solution, more relevant recommendations, etc.)
- [ ] Complementarity with other datasets ('missing piece of the puzzle')
- [ ] Other (please specify):

Q7. How large is the dataset that you obtained from your main (partnered) data provider for use in Data Pitch? *

Number of entries/observations/records *

Size in Gigabytes *

Q8. What is(are) the primary unit(s) of observation (e.g. customers, card transactions, patent filings, images etc.)? *

Q9. How large is the dataset that your solution uses in total? (I.e. any dataset(s) provided for Data Pitch + any dataset(s) you collected yourself or obtained from other data providers outside Data Pitch) *

Number of entries/observations/records *

Size in Gigabytes *

Q10. What is (are) the primary unit(s) of observation (e.g. customers, card transactions, patent filings, images etc.)? *

Q11. Please indicate the main type of data used in your solution, select all that apply. *

- [ ] Numeric
- [ ] Text
- [ ] Video
- [ ] Audio
- [ ] Geospatial
- [ ] Image/graphics
- [ ] Other (please specify):

Q12. How is this data stored? *

- [ ] Semantic databases (i.e. RDF triples)
- [ ] Document oriented databases
- [ ] Relational databases
- [ ] Files
- [ ] Other (please specify):

Q13. Where is the data stored? *

- [ ] Data Provider's infrastructure
- [ ] Commercial cloud paid for by the data provider
- [ ] Your own infrastructure
- [ ] Commercial Cloud paid for by your business
- [ ] Other (please specify):

Q14. How much control do you have over the data when building your solution? *

- [ ] Full control (e.g. full copy of data freely available to me)

- [ ] Partial control (e.g. data called through API when required)
- [ ] No control (e.g. I send my algorithms to the Data provider and get the results)
- [ ] Other (please specify):

Q15. What is the update frequency (periodicity) of the data used in your solution? *

	Static data (not updated)	Occasional data (irregularly updated)	Live data (regularly updated)	Real time data (regularly updated with high frequency)	Not applicable
During the acceleration period	☐	☐	☐	☐	☐
Mature commercial solution	☐	☐	☐	☐	☐

Q16. What resources/capabilities did Data Pitch funding enable you to acquire (rank in order of importance with 1 being most important ; answer N/A if Data Pitch funding was not used for a particular category)? *

	Ranking
Subject matter/domain knowledge	
Business management skills	
Software development skills	
Data science/machine learning skills	
Other IT skills	
ICT infrastructure/hardware	
Marketing/sales skills	

Please specify any other important resources/capabilities and their relative ranks

Q17. How closely did you interact with your partnered Data Provider? (1 signifies 'No interaction' and 5 signifies 'Very close interaction')

Q18. For each of the below, please rate the difficulty you had during the acceleration period:(1 signifying easiest, 5 signifying hardest)

Data Access

Data Engineering

Building the solution

Q19. What is the primary technical objective of your solution? *

- Combine and correlate different datasets
- Identify patterns
- Make predictions
- Enhance data quality
- Filter information
- Visualise information
- Create a user interface for data access

Q20. Does your solution use Machine Learning?

- Yes
- No

Q21. Which methods are used in your solution? Please select all that apply. *

- Regression algorithms (e.g. linear regression, logistic regression)
- Instance-based algorithms (e.g. k-NN, SVM)
- Decision tree algorithms (e.g. CART)
- Bayesian algorithms (e.g. naïve Bayes, Bayesian Network)

- [ ] Clustering algorithms (e.g. k-means, hierarchical clustering)
- [ ] Association rule learning algorithms (Apriori, ECLAT)
- [ ] Artificial neural network algorithms (e.g. MLP)
- [ ] Deep learning algorithms (e.g. CNN, RNN, DBN)
- [ ] Reinforcement learning algorithms (e.g. Q-Learning)
- [ ] Ensemble algorithms (e.g. Random Forest, GBM)
- [ ] Other (please specify):

Q22. How would you categorise your solution? Descriptive Analytics, which use data aggregation and data mining to provide insight into the past and future: "What has happened?" Predictive Analytics, which use statistical models and forecasts techniques to understand the future and answer: "What could happen?" Prescriptive Analytics, which use optimisation and simulation algorithms to advise on possible outcomes and answer: "What should we do?" *

- [ ] Descriptive
- [ ] Predictive
- [ ] Prescriptive
- [ ] Other (please specify):

Q23. How different is your solution now from the idea you had at the start of your involvement in Data Pitch? (An answer of 1 signifies your solution matches the initial proposal exactly; an answer of 10 signifies that the solution is completely different from the proposal).

Q24. Why is your solution different from the idea you had at the start of your involvement in Data Pitch? *

- [ ] Technical reasons (original solution was too difficult to implement during the acceleration period)
- [ ] Business reasons (changes to the solution resulting in a better/more marketable product)

- [ ] Would prefer not to say
- [ ] Other (please specify):

Q25. Who do you envisage as the primary customers for your solution? *

Q26. What distribution model do you envisage for your solution? *

- [ ] SaaS (Software as a Service)
- [ ] On-premise deployment
- [ ] Ad-hoc model (e.g. consultancy with engagement on a case-by-case basis)
- [ ] Would prefer not to say
- [ ] Other (please specify):

Q27. Are you considering releasing your solution as Open Source? *

- [ ] Yes, fully
- [ ] Yes, parts of it
- [ ] No
- [ ] Would prefer not to say

Q28. How do you see the scalability of your solution over the next three years? (An answer of 1 signifies 'limited growth', an answer of 5 signifies 'significant growth').

Core solution (markets & customers as currently identified)

Future adaptations of the core solution (new markets/application areas/customer groups)

Q29. How much effort do you foresee this scalability requiring?

	Minimum effort (e.g. only additional hosting space and computational power)	Medium effort (e.g. minimal software changes)	High effort (e.g. major software changes)
Core solution	☐	☐	☐
Future adaptations of the core solution	☐	☐	☐

Q30. On a scale to 1-10, how unique is the product that your solution provides to your customers? Are there similar types of products out there, or is this a one-of-a-kind? (An answer of 1 signifies the product is 'not at all unique' and an answer of 10 signifies the product is 'completely unique').

Q31. On a scale of 1-5, how innovative is your solution? (An answer of 1 signifies low innovation, an answer of 5 signifies high innovation).

Q30. Any other comments or remarks?

A3.5 Data Pitch evaluation – unsuccessful applicants survey

In the summer of 2017 and/or 2018, you applied to the Data Pitch programme. London Economics has been commissioned by the Data Pitch Consortium to carry out an independent assessment of this programme.

This survey will collect information on your interaction with the Data Pitch programme and your company's performance since applying to the programme. The survey will take approximately 15 minutes to complete.

If you have any questions, please contact Moritz Godel, T +44 (0)20 3701 7708, mgodel@londoneconomics.co.uk.

Q1. What is the name of your company? *

Q2. When did your company apply to Data Pitch? *

- ☐ Summer 2017
- ☐ Summer 2018
- ☐ Both

Q3. Is the company with which you applied to Data Pitch still active? *

- ☐ Yes
- ☐ No
- ☐ Prefer not to say

Q4. Approximately, how many people currently work at your company? *

- ☐ 1
- ☐ 2-5
- ☐ 6-10
- ☐ 11-50
- ☐ 51-250
- ☐ > 250

Q5. Approximately, what is your current annualised revenue? *

- ☐ €0 (pre-revenue)
- ☐ €1-€100,000
- ☐ €100,001-€500,000
- ☐ €500,001-€1 million
- ☐ €1 million-€2 million
- ☐ €2 million-€5 million
- ☐ €5 million-€10 million
- ☐ €10 million-€25 million
- ☐ €25 million-€50 million
- ☐ > €50 million

[Only for those who mentioned applying to Data Pitch twice] You mentioned that you applied to Data Pitch twice. In answering the following questions, please think of the last time you applied to Data Pitch.

Q6. Why did you apply to Data Pitch? Select all that apply. *

- ☐ Funding
- ☐ Access to data
- ☐ Other (please specify):

Q7. How easy was it to complete and submit your application? *

	1	2	3	4	5	
Very easy	☐	☐	☐	☐	☐	Very difficult

Q8. How transparent was the decision-making process around your application? *

	1	2	3	4	5	
Very transparent	☐	☐	☐	☐	☐	Very opaque

Q9. Do you have any other comments on the application process for Data Pitch?

Q10. By how much has the number of employees in your company changed in the first 12 months after applying to Data Pitch? *

- [ ] Increased by more than 10 employees
- [ ] Increased by 6 to 10 employees
- [ ] Increased by 1 to 5 employees
- [ ] No change
- [ ] Decreased by 1 to 5 employees
- [ ] Decreased by 6 to 10 employees
- [ ] Decreased by more than 10 employees

Q11. By how much did your annualised revenue change in the first 12 months after applying to Data Pitch compared to before applying? *

- [ ] Increased by more than €1 million
- [ ] Increased by between €500,001 and €1 million
- [ ] Increased by between €100,001 and €500,000
- [ ] Increased by between €1 and €100,000
- [ ] No change
- [ ] Decreased by between €1 and €100,000
- [ ] Decreased by between €100,001 and€500,000
- [ ] Decreased by between €500,001 and €1 million
- [ ] Decreased by more than €1 million

Q12. Generally, how challenging is access to external data sources for your company? *

	1	2	3	4	5	
Very easy	☐	☐	☐	☐	☐	Very challenging

Q13. How much external funding did you receive in the first 12 months after applying to Data Pitch? *

- [ ] We received no external funding after applying to Data Pitch
- [ ] Less than €10,000
- [ ] Between €10,000 and €50,000
- [ ] Between €50,000 and €100,000
- [ ] Between €100,000 and €250,000
- [ ] Between €250,000 and €500,000
- [ ] More than €500,000

Q14. What type of funding did you receive? Select all that apply. *

- [ ] Seed investment
- [ ] Angel investment
- [ ] Series A round investment
- [ ] Series B round investment
- [ ] Series C round investment
- [ ] Mezzanine investment
- [ ] Other (please specify):

Q15. Where did the funding originate from? Select all that apply. *

- [ ] From within the country we are based in
- [ ] From within the EU, but not the country we are based in (including EU institutions)
- [ ] From outside of the EU

Q16. Do you have any other comments or remarks?

Q17. Can we contact you with further questions? *

- [ ] Yes
- [ ] No

If you agree that we can contact you, we will need to collect some personal information. Our full privacy policy is available here.

Q18. Please provide your name and contact details.

Name:	
E-mail address:	
Telephone number:	
Any other contact details:	

Annex 4 Note on quantitative evaluation

As part of the impact assessment, London Economics investigated the possibility of providing quantitative estimates of the effect of Data Pitch on the participating companies and the wider economy. Following our initial proposal, we explored the use of a Fuzzy Regression Discontinuity Design (FRDD) to provide impact estimates[74].

A4.1 The Fuzzy Regression Discontinuity Design

The impact of the Data Pitch programme can be assessed by comparing the outcomes for companies that took part in the programme (successful applicants) with those of similar companies that did not take part. In an ideal world, one would take a collection of companies, randomly divide them into two groups and administer the programme to one of the groups and not the other. Because of the randomisation, the two groups would be comparable and the comparison between the two groups provides an estimate of the programme's impact.

However, in Data Pitch companies are not randomly assigned to take part. There is an application phase that (non-randomly) selects successful applicants. However, the selection of successful applicants was based on an observable measure; all applications were scored and selected for an interview if the score exceeded a defined cut-off value[75]. If this score was the only determinant of being selected, and if the cut-off had been applied rigidly, then a (sharp) Regression Discontinuity Design (RDD) could have been applied.

RDD works on the assumption that applications with similar scores should be of a similar quality, and therefore should be equally likely to succeed under the same circumstances. If this is true, then we can compare successful applicants that scored just above the cut-off with unsuccessful applicants scoring just below the cut-off. These two groups of companies should be comparable in quality, but one group took part in the programme and the other did not.

In effect, RDD creates quasi-experimental conditions where one is comparing two groups that differ only in the fact that one group is "treated" and the other is not. The difference between these groups, as in an experiment, provides an estimate of the effect of the programme.

In Data Pitch, the score was not the only determinant of whether an applicant would be successful. The application phase was followed by an interview phase that further screened out applicants. Furthermore, the cut-off was not rigidly applied, to ensure the best use of the programme's available budget. A **sharp** Regression Discontinuity Design is therefore not possible, but a **fuzzy** Regression Discontinuity Design (FRDD) may still be possible.

The FRDD does not depend on a rigidly applied score and can be used as long as the **probability** of being selected for the programme significantly increases at the cut-off point. If that is the case, econometric techniques exist that can account for applications that passed the cut-off but were unsuccessful in the end.

[74] FRDD is a state-of-the-art method that has been used in similar settings, notably by Bone et al. (2019).

[75] Cohort 1 applications were scored out of 100 with the cut-off at 60. Cohort 2 applicants were scored out of 5 with the cut-off at 3.

A4.2 Can FRDD be applied to Data Pitch?

A4.2.1 Evidence of an exploitable discontinuity in application scores

To determine whether it is possible to implement FRDD for the Data Pitch evaluation, we first looked at whether there is indeed a discontinuity at the cut-off in the probability of being selected, i.e. whether the probability of selection "jumps" at this point.[76]

The graph below shows the average probability of being selected for Data Pitch plotted against the application score[77]. The red line shows a linear regression through the dots, allowing the line to be different on either side of the cut-off value. The graph clearly shows that:

1) there is a positive relationship between the application score and the ultimate probability of being selected for the Data Pitch programme, and:
2) there is a clear increase in the probability of being selected at the cut-off value of 60[78], as shown by the discontinuity in the regression line.

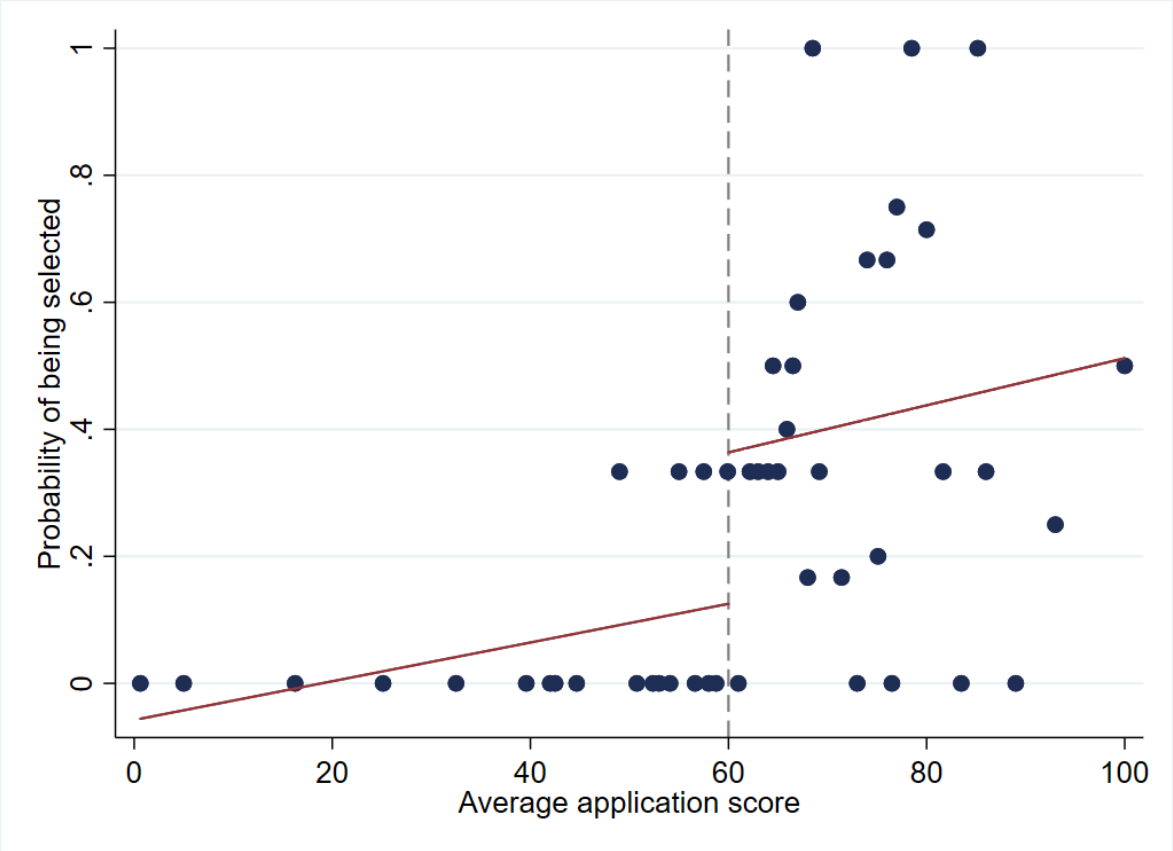
Further testing shows that the probability of being selected for the programme increased by about 29 percentage points at the cut-off value. This jump is highly statistically significant ($p < 0.01$). In principle, therefore, FRDD could be applied.

[76] Normally, one would also check for the possibility that scores were manipulated to favour certain companies that should not be selected based on the scoring criteria. This would undermine the assumption that applications with similar scores are of similar quality. Data Pitch's review process accounted for potential conflict of interests by ensuring that applications were reviewed by a reviewer without a relation with the reviewed company. Furthermore, the reviewers were not disclosed to the applicants. Therefore, there was no incentive for reviewers to manipulate scores.

[77] The dots represent the probability of being selected per bin of two. In other words, it shows the probability for applications with scores (where available) 0-2, 2-4, 4-6, etc.

[78] The graph shows grouped data for both cohorts. The average scores for cohort 2 were multiplied by 20 to get a score out of 100 with the cut-off at 60.

Figure 43 Probability of being selected for the Data Pitch Programme



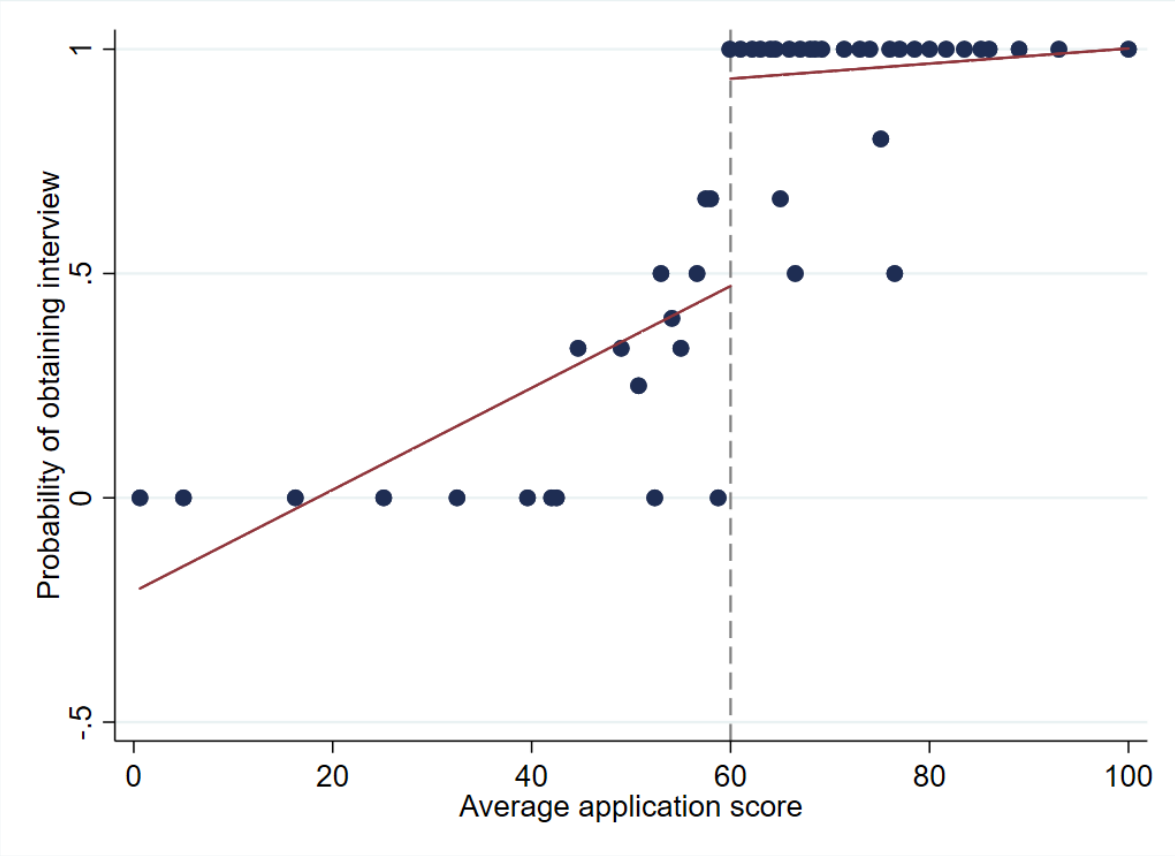
Notes: Based on 165 applications. The scores of two applications were not known at time of writing. Cohorts are grouped.

Source: Data Pitch materials

A similar graph can be used to show that the cut-off value of 60 to obtain an interview was not rigidly applied. The graph below shows a similar graph as Figure 43 but, instead of showing the probability of being selected for the programme, it shows the probability of obtaining an interview.

The graph shows that the probability of obtaining an interview did jump by more than 50 percentage points at the cut-off of 60. However, some companies with a score lower than 60 were invited for an interview, and some companies with a score higher than 60 were not.

Figure 44 Probability of obtaining an interview in the application process



Notes: Based on 165 applications. The scores of two applications were not known at time of writing. Cohorts are grouped.

Source: Data Pitch materials

A4.2.2 Availability of outcome measures

Having established that a quasi-experimental impact assessment is feasible in principle, we considered suitable outcome measures for the analysis. Suitable outcomes measures have to satisfy two conditions:

1) they have to capture change over time; and,
2) they have to be available for both successful and unsuccessful applicants.

Examples could be survival of companies, increases in employment or funding attracted by companies since applying to Data Pitch. **Our conclusion at this stage is that the necessary data is not available:**

- **Survival rates:** The end of the Data Pitch programme is too recent, and the programme duration too short, for survival rates to be useful as outcome measure:
  - No impact on survival rates can be observed for cohort 2 yet.
  - Most unsuccessful applicants still exist. Of all unsuccessful applicants, we identified nine as inactive with a further 4 for which the status could not be discerned.
- **Employment:** While employment data is available for the successful applicants, this data could not be obtained for unsuccessful applicants. We investigated the use of external source for employment data for unsuccessful applicants (Orbis and Crunchbase), but here data is scarce. We could only locate employment figures for around 30% of all applicants.

- **Funding:** Similarly, funding data for unsuccessful applicants (Crunchbase) is sparse and incomplete.

We considered other outcome measures such as advancement in technology readiness and business development stage, but were not able to overcome the data limitations

LE
London
Economics

Somerset House, New Wing, Strand,
London, WC2R 1LA, United Kingdom
info@londoneconomics.co.uk
londoneconomics.co.uk
@LondonEconomics
+44 (0)20 3701 7700