Review of the UK business to business data assurance market

Final report

July 2021

ODI Foreword

To create a world where data works for everyone, and unlock value for the UK economy, we need trustworthy data to flow in well-governed ways around the data ecosystem. Data sharing is restricted when governments and companies that could share or reuse data are unable to assess the trustworthiness of datasets or other organisations. This can reduce their willingness to provide access to data they steward and to use data from third parties, leading to missed opportunities for the use of data and duplicated data collection effort.

The role of trust and assurance in increasing data flows is an area that has been at the centre of ODI work for a number of years. In 2019 we commissioned public innovation consultancy, Oxford Insights, to undertake exploratory work around certifications for data trusts. This examined what it would mean to give an organisation stewarding data on behalf of others an official stamp or seal of approval. We built on this through work to evaluate the economic impact of trust in data ecosystems and found that trust, ethics and openness around data directly link to increased data flows and value creation, up to 2.5% of GDP. In parallel we researched the role of certification and audit in building trust and trustworthiness in data and data practices. We found that trust and trustworthiness are highly context dependent and that third-party assessments are useful, but only to a degree and within certain contexts.

Data assurance activities have a role to play by helping organisations assess, build and demonstrate both trust in and the trustworthiness of data and data practices. Organisations and individuals that access, use and share data need to provide assurance, or be confident, that data is fit for purpose and trustworthy for use, and that the purposes of data collection are ethical and equitable. Policy makers and organisations that offer data assurance services products and services need to understand where they can best support trustworthy data sharing. Identifying where they can add value requires an understanding of the current market for UK data assurance; including where the market is flooded, and areas where there are gaps.

We are therefore delighted to introduce this work by economics consultancy Frontier Economics, commissioned as part of a wider programme of work on data assurance, aimed at improving the data practices of organisations so that they can build and manage adequate data infrastructure and data use, which took on exploration of this topic.

This research reveals a potentially significant market opportunity for organisations that are concerned with increasing trust in data and data practices. We expect to see high growth in years to come, as more organisations grapple with demand to build confidence with customers, shareholders and the public. As we've recently highlighted in our response to the government's consultation on restoring trust in audit and corporate governance, data assurance practices are at the heart of building confidence in the way that UK companies are run and scrutinised. At the ODI, we're looking forward to working with the market to develop practices, products and services in this increasingly significant area.

Deborah Yates, Programme Lead - Data Assurance

Contents

Data assurance is about building confidence in data and the processes that support creation, storage and usage of that data

Assurance is about building confidence and demonstrating trustworthiness to others via evidence

Data assurance services can enable organisations to:
- understand risks and opportunities in the data they create and share themselves;
- trust the data they access and reuse from third parties
- share data with others who they are confident will avoid harmful impacts while using it.

Data assurance can be a process that organisations undertake on an ongoing basis. These processes may occur continuously in real time or be centred around certain events such as data audits.

Data assurance can play a role at multiple stages of the data value chain from verification of data collection practices to building trust which facilitates more appropriate data curation

Data value chain

| Data collection/ sourcing | Data harmonisation | Data sharing/ publishing | Data interlinking | Data exploitation | Data curation |

Source: adapted from Attard et al. (2016) and Attard et al. (2017)

Frontier and glass.ai were commissioned to review the UK data assurance market

We have developed a four stage methodology featuring an innovative form of data collection and analysis:

Step 1	Step 2	Step 3	Step 4
Desk review	Gather quantitative data	Analyse quantitative data	Reporting

1 What is the size of the market?

2 How mature is the market?

3 What types of firms are active in this market and where are they based?

4 What are the types of products and services offered by the firms?

We addressed four research questions
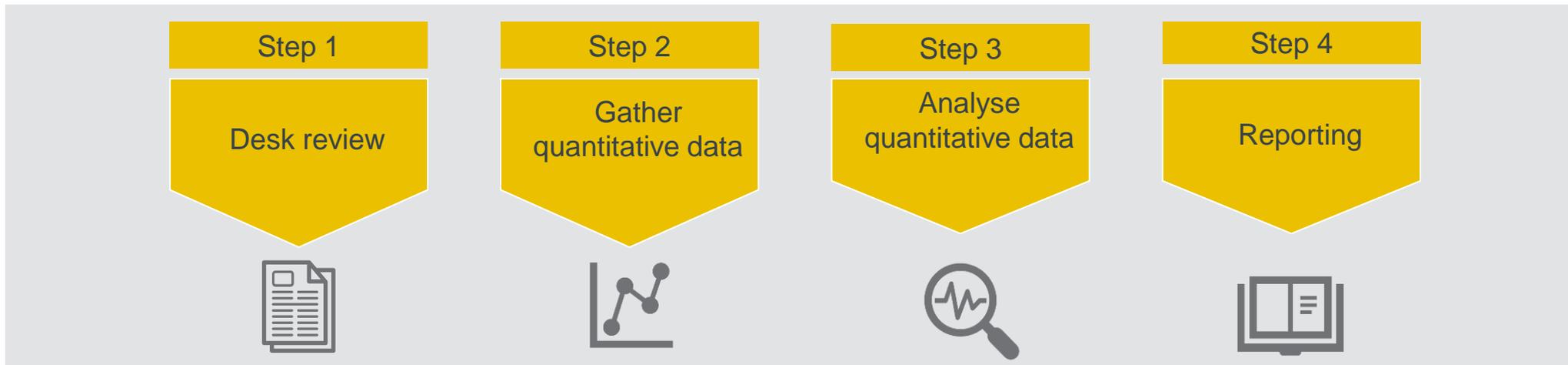
# We have four key takeaways from our work

SIZE OF MARKET

There is a considerable amount of economic activity in the data assurance space in the UK.

Employment in firms offering data assurance firms is spread across a wide number of firms of different sizes

MATURITY OF MARKET

The majority of firms we identified who provide data assurance services were incorporated in the last 10 years.

The average birth rate amongst data assurance firms is almost 12% per annum.

TYPES OF FIRMS AND LOCATIONS

In keeping with other emerging sectors, it is hard to identify a specific data assurance market. Data assurance activities may in several cases have become part of the offer of firms active across different traditional sectors

TYPES OF PRODUCTS AND SERVICES
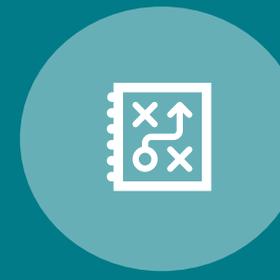
A range of data assurance products and services are available. These offerings may be aimed at different types of users with different needs. Several keywords did not match against any firms. This could provide tentative evidence of gaps in the market.

Most firms engaged in data assurance activities have fewer than 10 employees

SIZE OF MARKET

In total, we identified 890 firms engaged in data assurance activities (for simplicity we refer to these as "data assurance firms" throughout, even though in some cases their commercial activity will not be entirely data assurance related). We have conservatively estimated that approximately 30,000 staff are involved in the delivery of data assurance activities. This assumes that only a small proportion of total employment in the largest firms we identified is related to data assurance.

The majority of identified firms have a relatively small number of employees each

- 56% of the firms who are engaged in data assurance activities had 10 or fewer employees and 87% have fewer than 50 employees. We identified seven firms who each employ over 10,000 staff (these are primarily large audit / accounting firms). The majority of these employees in the largest firms will not be directly involved in data assurance activities.

KEY TAKEAWAY

There is a considerable amount of economic activity in the data assurance space in the UK. Employment in firms offering data assurance is spread across a wide number of firms of different sizes

Breakdown of data assurance firms by employee size band

Source: Frontier analysis of data collected by glass.ai

Note: The total employment figure is based on the number of employees who have listed themselves as an employee of a firm on LinkedIn. We have assumed that only 1% of employees in firms with >1000 employees are engaged in data assurance activities. The firm level employment band data is a separate variable. These two separate measures of employment will be highly correlated but will not always be completely consistent.

The market is growing at a rate which is broadly consistent with the wider economy

MATURITY OF MARKET

**Between 2014-2020 an average of 44 new firms entered the data assurance market each year**

- The average birth rate amongst data assurance firms is 11.7% per annum.
- In total 58% of the data assurance firms that we identified were incorporated in the last ten years and 90% were incorporated after 2000.

KEY TAKEAWAY

The majority of firms we identified who provide data assurance services were incorporated in the last 10 years.

Market entry over time
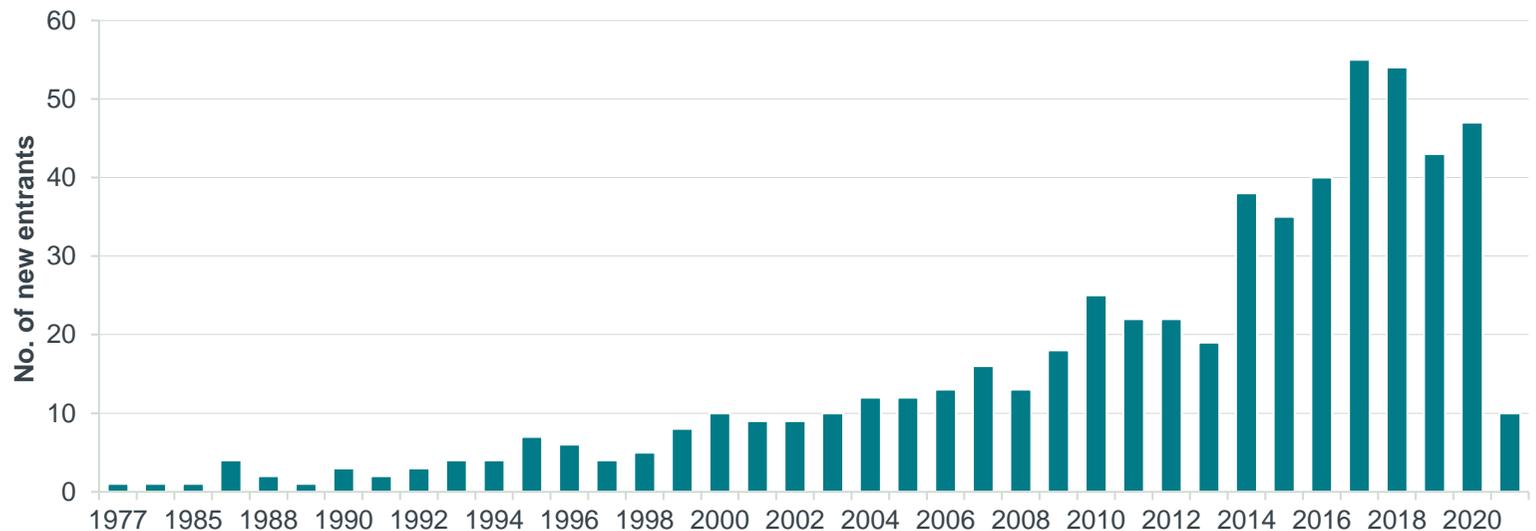
Source: Frontier analysis of data collected by glass.ai

Firms who engage in data assurance activities operate across a range of traditional sectors

TYPES OF FIRMS AND LOCATIONS

Data assurance firms specialise in a variety of technology, finance and professional service areas

- Using the glass.ai classification system (see Annex for full breakdown) we found the most common sectors included Information Technology and Services, Consulting, Computer Software, and Computer Networking and Security. This illustrates the wide range of sectors that data assurance companies are active in. It also reinforces the point that we cannot isolate data assurance activities within a single traditional sector.

KEY TAKEAWAY

In keeping with other emerging sectors, it is hard to identify a specific data assurance market. Data assurance activities may in several cases have become part of the offer of firms active across different traditional sectors

Top ten glass.ai sub-sectors: percentages

Information Technology and Services, 44%

Consulting, 14%

Computer Software, 13%

Computer Networking and Security, 8%

Others, 5%

Security and Investigations, 3%

General Financial Services, 3%

Information Services, 4%

Accounting, 2%

Law Practi… and Servi…

Oil and Ene… 1%
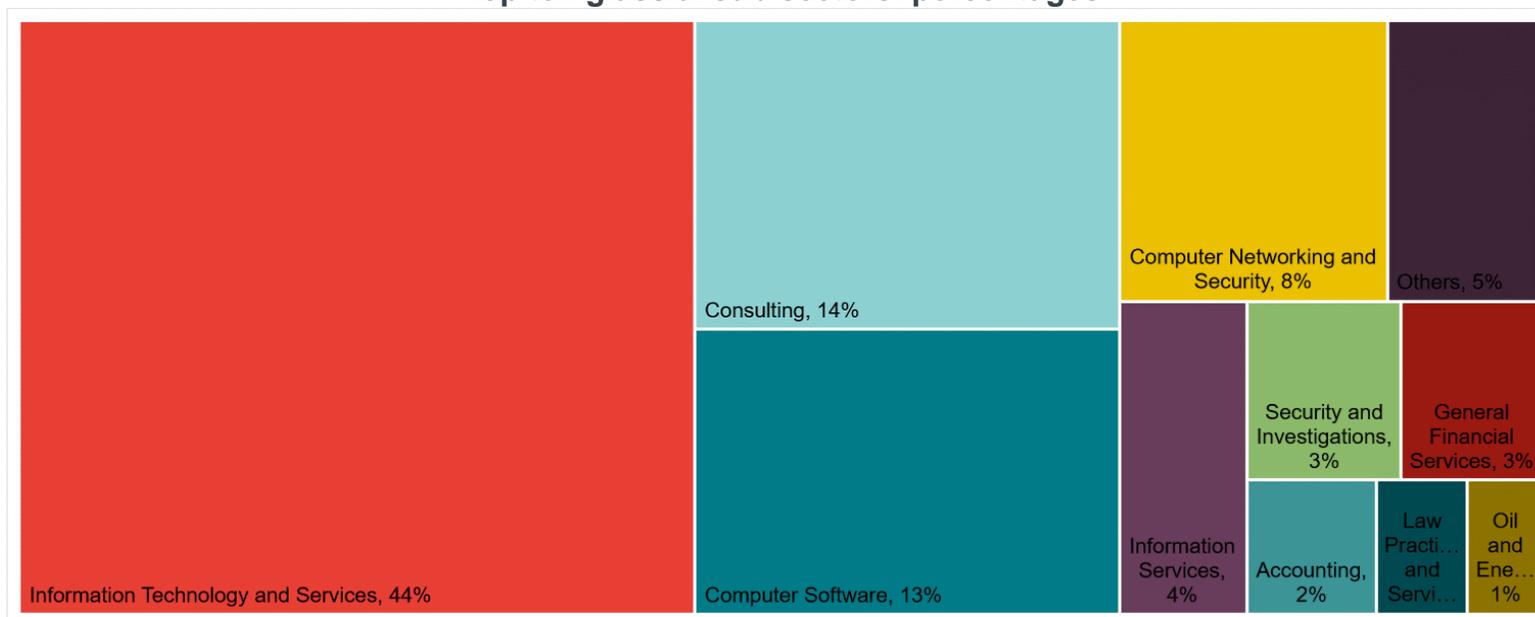
Source: Frontier analysis of data collected by glass.ai

Firms engaging in data assurance activities are overrepresented in London relative to share of all businesses in the region.

TYPES OF FIRMS AND LOCATIONS

KEY TAKEAWAY

Firms engaged in data assurance activities are more common in certain geographic areas than others. There is a clear clustering of firms around London and the South East. There are pockets of data assurance activity in all regions but this is limited to urban centres generally.

Firms engaged in data assurance activities are based in clusters across the UK

We see that there is a clear clustering of firms around urban centres especially around London which contains 40% of the firms in our sample. This is more than twice the proportion of all UK firms who are based in London

Geographic distribution of UK firms engaged in data assurance activities

Region
- East Midlands
- East of England
- Isle of Man
- London
- North East
- North West
- Northern Ireland
- Scotland
- South East
- South West
- Wales
- West Midlands
- Yorkshire and The Humber
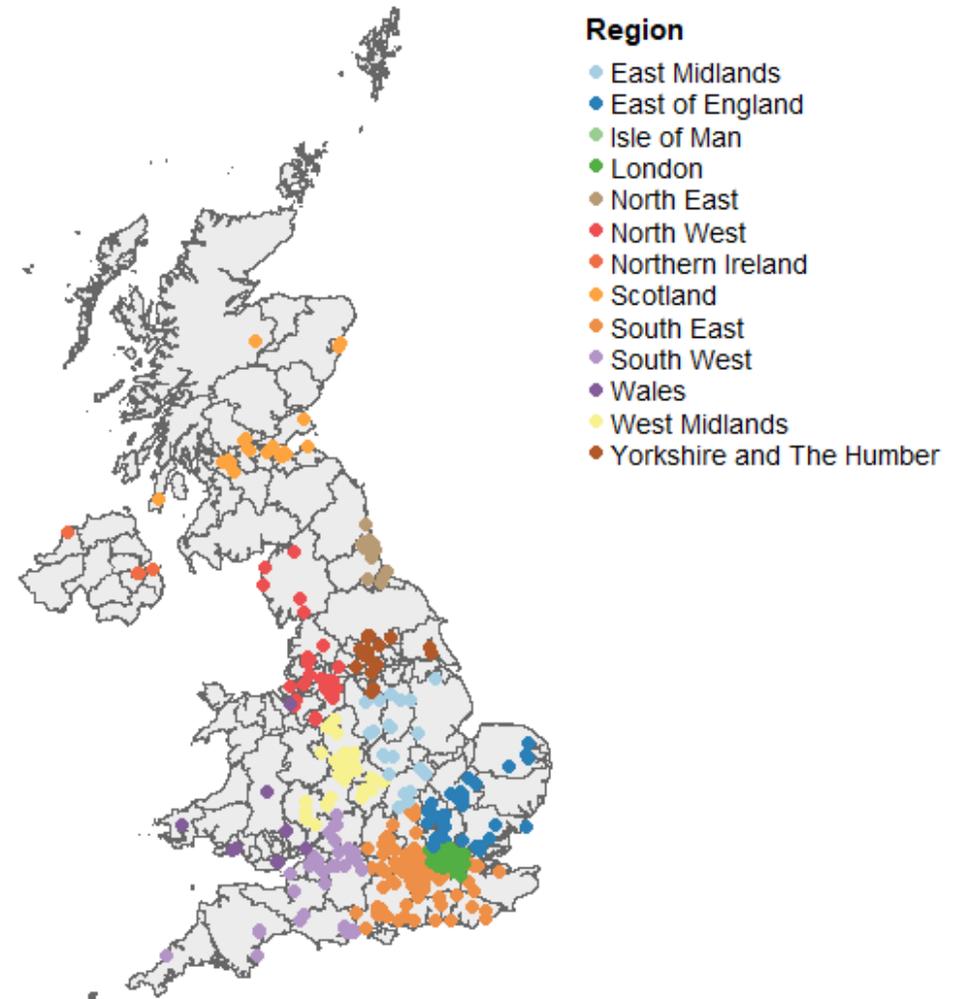
Source: Frontier analysis of data collected by glass.ai

Data management and data governance were the two most identified search terms

TYPES OF PRODUCTS AND SERVICES

KEY TAKEAWAY

A range of data assurance products and services are available. These offerings may be aimed at different types of users who have different needs. Several search terms did not match against any firms. This could provide tentative evidence of gaps in the market.

Most of the data assurance firms are identified through a subset of topic matches

- Overall 52 of our 75 search terms identified at least one firm that was included in our final list. 48% of the firms we identified were linked to one of the top 5 most frequent topics.

- Only 8 out of the 890 firms identified were linked with the specific search term 'data assurance'.

Our deep-dive into large accounting firms' data assurance offerings revealed a number of service offering categories:
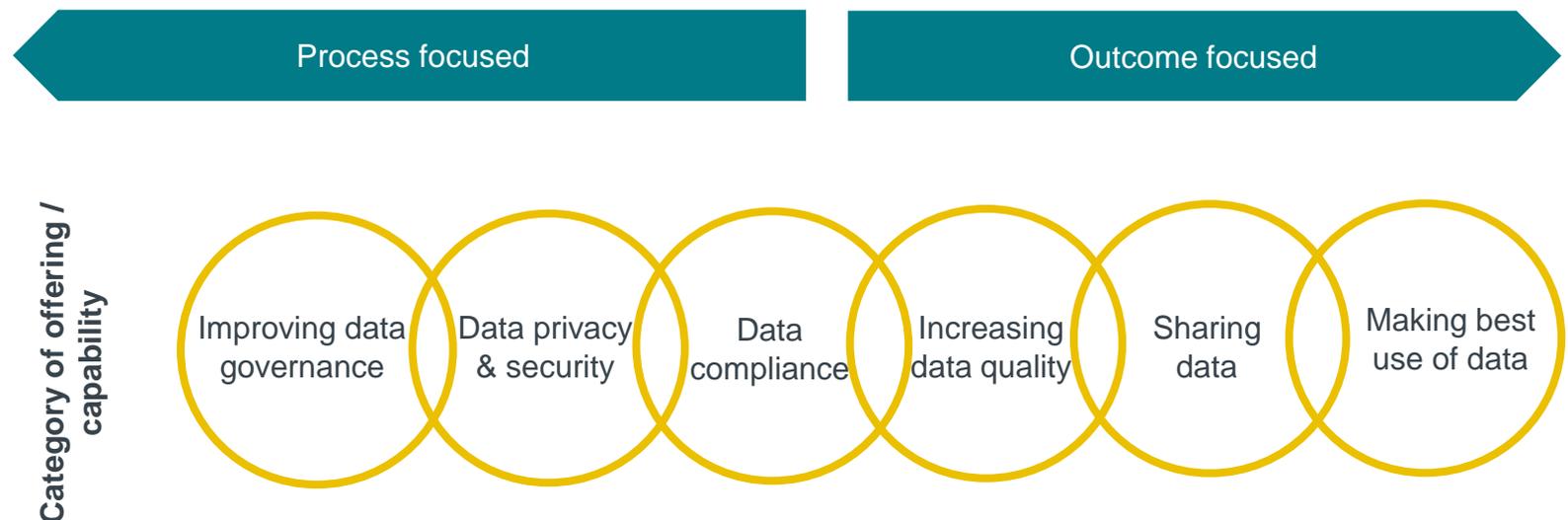
Process focused	Outcome focused

Category of offering / capability

- Improving data governance
- Data privacy & security
- Data compliance
- Increasing data quality
- Sharing data
- Making best use of data

frontier economics

The Open Data Institute have commissioned Frontier to examine the UK market for business-to-business data assurance services

The mission of the Open Data Institute (ODI) is to work with companies and governments to build an open, trustworthy data ecosystem

The ODI have started a programme of research on data assurance

- ODI are looking into the role of data assurance. ODI has previously worked with companies and governments to both help them assure data, and be assured by the quality of the data provided by others

- This forthcoming programme of work builds on previous work carried out and commissioned by the ODI. The ODI has carried out extensive research into how data is collected, managed and used. Previous ODI research has been focused on exploring which mechanisms are likely to have the most impact in improving trust between organisations around data. Some of the ODI work in this area investigates:
  - the levers through which trust in data ecosystems can be enhanced;
  - the ways in which organisations can demonstrate trustworthiness when sharing data; and
  - the value of sharing data to build trust and trustworthiness.

Scope of this report

Frontier and glass.ai were asked to carry out:

"a market review to improve understanding of the current UK data assurance market"

- Our work helps inform decisions on where the ODI and other actors can best support adoption products that will lead to greater trust and data sharing. The work will also serve as a benchmark for later years.

We have focused on four specific research questions throughout our work

What is the size of market?

- How many firms are active in this space currently in the UK?
- What is the size distribution of firms currently active in this market in terms of revenue and number of employees?
- What is the aggregate size of this market in terms of employees and revenue?

How mature is the market?

- What is the distribution of incorporation dates amongst included firms?
- How do recent market joiners compare to more established firms?

In what sectors and areas are firms based?

- Across what sectors of the economy are data assurance firms most often categorised?
- Where in the UK do data assurance firms have their HQs?

What are the types of products and services offered?

- What types of offering exist in the market?
- What are the typical business models adopted by data assurance firms?

Data assurance services can help to build confidence in data and data practices

Assurance is about building confidence and demonstrating trustworthiness to others via evidence.

Data assurance is the process by which organisations assess, and demonstrate trustworthiness in their own data and data practices as well as the data and data practices of others

For the purposes of this work we are defining **data assurance** as the process by which:

"*organisations **assess**, and **demonstrate** trustworthiness in their own data and data practices as well as the data and data practices of others. This in turn can help **build** trust and confidence in data ecosystems.*"

Organisations across the economy may undertake internal data assurance activities and they may also work with commercial data assurance specialists. These specialist organisations provide business-to-business offerings that allow organisations to:

- understand risks and opportunities in the data they create, use and share themselves,
- to trust that the data they access and reuse from third parties
- to share data with confidence that those they are sharing with are set up to use it while avoiding harmful impacts.

Data assurance processes are undertaken on an ongoing basis

Assurance is an ongoing process which informs decision making in organisations

**Data assurance will be a process that organisations undertake on an ongoing basis. These processes may occur in real time or be centred around certain events such as data audits**

- We are considering data assurance an ongoing process which allows organisations to take informed decisions about the data they hold currently or will collect in the future. **Firms engaging in data assurance activities may provide guidance, expertise and technical solutions as part of an overall offering.** This enables their clients to make informed decisions with regards to the responsible usage of data to achieve their strategic goals.

- Data assurance tools can help inform these decisions but some level of judgement will still be required and application to the relevant context at hand will be needed. **The precise data assurance services that an industry or specific firm requires will vary depending on the type and volume of data collected, its underlying characteristics and the objectives of the organisation using the data assurance service.** Therefore a one-size-fits all approach to data assurance or a "tick-box" exercise is unlikely to be successful.

Data assurance services can be used by data holders and data users

Organisations who hold data need to be able to provide assurance that the data is fit for use (and possible onward sharing) by others. Data holders' assurance activities could include implementing transparent quality control procedures and providing background information on how information has been collected and what rights to reuse the data have been provided, through clear licensing.

Data holder
(individuals and organisations that can created or collected data)

Data users
(individuals and organisations that can benefit from accessing data)

Some types of data users, in particular businesses seeking access to data, need to be able to assure others that they are trustworthy in their use of data. For example, data users may engage with data assurance providers to demonstrate how their data storage processes are robust and meet industry standards to allay fears of data misuse and/or leakage.

Firms may engage with data assurance services to avoid negative outcomes

① AVOID NEGATIVE OUTCOMES

Firms may be motivated by a desire to avoid negative outcomes associated with **violating existing data rules and regulations**

- The **rollout of economy-wide regulations such as the General Data Protection Regulation (GDPR) has raised awareness about the importance of appropriate data governance and privacy procedures in relation to personal data**. Specific industries will also have their own rules, codes of conduct and/or best practices which individual organisations will want to adhere to.

- **Data breaches or other negative shocks can lead to fines and temporary or permanent losses of revenue or profitability**. We have seen several of these occurring in recent years.

Role of data assurance

Data assurance offerings such as "certification" and "accreditation" can help to ensure that organisations have appropriate processes in place to comply with existing rules and regulations (Oxford Insights, 2019).

- Certification refers to evaluating processes or systems against certain standards.
- Accreditation assess the competence of an organisation to perform certain data related tasks

Firms may engage with data assurance services to maximise value from data assets

(2) MAXIMISE THE VALUE FROM DATA ASSETS

Firms increasingly want to make the most of the data they hold and ensure that business data becomes a key organisational asset.

- As data has become abundant in the modern world an **increasing number of organisations are realising that data they collect and/or hold can drive productivity and inclusion** either via improved internal decision making or external usage.

Role of data assurance

Data assurance offerings can help to

- boost internal trust in an organisation's data and therefore lead to greater acceptance of evidence-based decision-making in relation to commercial operations.

- increase consumers' trust in a business which in turn can lead to greater brand value or profitability.

Data assurance services are one avenue by which data sharing can be encouraged which leads to better societal and economic outcomes

Data assurance can lead to greater data sharing

Increasing the proportion of decisions taken based on data can generate significant benefits for all of society.

- **Data assurance processes (in conjunction with other factors) can enable greater data sharing** (to underpin these decisions) help provide confidence that data informed decisions are reliable.

- **For example**, a public sector health data holder may be willing in principle to share the data they have collected with academics who are seeking to carry out a piece of socially beneficial research. However, this sharing may not happen if the data holder has concerns about the processes and infrastructure that the researchers have in place to store sensitive data. Data assurance services can help to overcome this barrier:

The theory of change of data assurance activities helps to lay out the logical flow of how they can lead to societal and economic benefits

Activities: drivers of trust	Outputs: changes in trust	Outcomes: changes in data usage	Impacts: societal and economic benefits
Data assurance activities allow organisations to assess, and demonstrate trustworthiness in their own data and data practices as well as the data and data practices of others. A potential data user (such as the academic in our example above) can use data assurance to demonstrate the quality of their data practices.	Increased data assurance activities can (in conjunction with other factors) help build trust and confidence in data ecosystems.

A data holder (such as the public sector organisation in our example above) may share data with greater confidence if they see potential users demonstrating their competence via data assurance services. | Greater levels of trust can lead to more data being shared and used. This in turn leads to better decisions actions and insights.

In our example above this could include more healthy data being shared and socially beneficially research being carried out to inform future healthcare design for example. | Consumer welfare

Efficiency

Gross value added

Wider societal benefits |

A data value chain can help to identify the stages at which value is created

A data value chain describes the data lifecycle steps which create value

Data value chains set out the interactions between different activities which can add value in the context of collection and usage of data. These value chains can provide a useful perspective on creating value on a digital, intangible product.

An example of a data value chain

| Data collection/ sourcing | Data harmonisation | Data sharing/ publishing | Data interlinking | Data exploitation | Data curation |

Value is created/ sourced by generating data via primary collection or acquisition

Data is then transformed into a more useful and standardised format to help in processing

This refined data can then be published and shared externally

In some cases users will then link this data with other sources of data to add additional value

Organisations may then use this aggregated and processed data as a means to inform organisational decisions and add value internally.

Finally, the value created from the data will depend on ongoing maintenance, updating and cleaning of data

Data assurance can play a role at any stage of this value chain

We can consider the potential role of data assurance at each stage of this value chain

- Data assurance can play a role at each part of the value chain from initial collection of data to final exploitation and usage of data either by the organisation who has collected the data or another organisation.
- We have set out example data assurance offerings below which are not intended to be exhaustive.

Data collection & sourcing	Data harmonisation	Data sharing / publishing	Data interlinking	Data exploitation	Data curation

Source: adapted from Attard et al. (2016) and Attard et al. (2017)
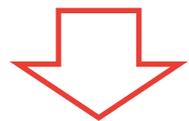
|					
Data assurance can help demonstrate that data collection processes are ethical and equitable	Data assurance offerings can help demonstrate that data and underlying data practices conforms to existing standards	Assurance plays a role in building trust that data is "fit for sharing"	Assurance can help build trust that data is consistent with other sources of data and confirm that linking does not create ethical issues.	Data assurance enables trust in the insight, decision making and actions taken from data.	Assurance offerings can demonstrate the curation activities are taking place.
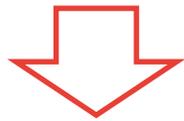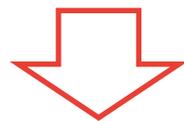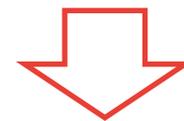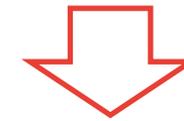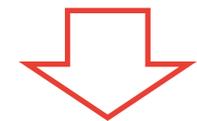
frontier economics

We have developed a four stage methodology to address the core research questions

Step 1	Step 2	Step 3	Step 4
Desk review	Gather quantitative data	Analyse quantitative data	Reporting

During the desk review stage we developed our understanding of data assurance and listed "key words" which informed our data collection

Step 1

Desk review

Step 1. Desk review

- We kicked-off the work by carrying out a rapid desk based review of the data assurance market. This built on the tools and resources identified by ODI as part of their previous work on building trust around data (2021 A). We carried out online keyword searches to understand the scope of existing work that has been carried out in this space and to identify examples of firms active in this space.

- The outputs of this step in the work were: (1) A shared understanding of what is meant by data assurance and how it relates to other interrelated topics (2) examples of firms active in the UK data governance market and (3) examples of data assurance products, services and tools that are currently in use.

- At the conclusion of this stage of work we proposed a list of search terms (topics and keywords) that were used to inform the subsequent quantitative data collection stage (see Annex for full list).

Our desk research suggested multiple categories of key words

AREA OF FOCUS

We are considering data assurance as an ongoing process which allows organisations to take informed decisions about the data they hold currently, will collect in the future, might access or share. The search terms we defined and subsequent filtering we have applied have been tailored with this focus in mind.

Out of scope

We are less interested in organisations who are offering standardised technology platforms which can be implemented en masse in a variety of settings.

In scope

firms who provide ongoing bespoke guidance, expertise along with technical solutions to support their clients to make informed decisions with regards to the responsible usage of data.

SEARCH TERM CATEGORIES

Overarching "foundational" terms

Characteristics of trustworthy data

Data assurance services

Data assurance products

Avoidance of negative outcomes

We employed an innovative web-crawling technique to identify firms active in this space

Step 2

Gather quantitative data

We employed a pragmatic approach that relied on an innovative form of web crawling which allowed us to gather as much information as possible about firms active in the supply side of this market within a short space of time.

1

The glass.ai engine has an ongoing process that regularly crawls the websites of over 2 million UK businesses. This is augmented with other related content about the businesses gathered from news, social media and official data sources like Companies House. The content is classified using proprietary artificial intelligence (AI) technology that understands written language. Specifically, the glass.ai engine is able to identify descriptions of the business activities that the company is involved in and apply these to further analysis.

2

After the list of keywords was agreed following input from ODI we used this AI capability to identify any web text that may suggest the companies are involved in data assurance-related activities. This process generated a bespoke dataset of UK data assurance companies and associated descriptions. As set out previously this search was informed by parameters and search terms from our desk research.

4

For each firm identified, we used company metrics sourced from Companies House and social media to populate a number of data fields such as the description of economic activity, industry sector, size of the companies (measured in terms of number of employees and turnover), data assurance topics identified, company location and age of company. To quality-assure the data, both manual and automatic quality assurance processes were applied to the data fields (see Annex).

3

In addition, we ran "deep crawls" of the larger accounting firms' websites, some of whom we knew to be active in the data assurance market, to identify the specific types of data assurance services they offer. The advantage of this approach was that although these services only form a small part of their offering (for example they may not appear in main descriptions of these businesses activities), the deep crawls ensured that these were picked up regardless of where they were described.

We analysed and visualised results to address each of our research questions

Step 3. Analyse quantitative data

▪ The fourth stage of the study involved analysing the quantitative data that has been collected via the innovative web scraping procedure we have described above. We used that dataset to answer our research questions of interest for this piece of work.

▪ Specifically, the data that we collected was used to provide nuanced insights across a number of areas that relate to the supply side of the data assurance market. This quantitative analysis provided granular information on: the number of firms in the market; the overall size of the market; the location of firms in the market; the traditional sector in which these firms operate.

▪ We have illustrated these findings in the remainder of this report.

This report summarises all of the work we have undertaken and the conclusions we have reached

Step 4. Reporting

- Staff at the ODI met and corresponded with Frontier Economics regularly in order to develop and steer this research project and provided feedback on an earlier draft of this report.

- In addition, to this final report we have also shared the underlying dataset with ODI which will help to inform their future activities in this space.

- All IP directly generated as a result of this work is the property of ODI.

frontier economics

Our first research question focuses on the size of the market for firms engaged in data assurance activities

SIZE OF MARKET

Quantifying the overall size of the data assurance market is an important first step in understanding the nature of provision.

Our first research question focuses on the overall scale of commercial activity in relation to data assurance in the UK economy.

Measuring the size of the data assurance market includes a number of specific sub-elements:

1 Number of firms engaged in data assurance activities active in the market

2 Number of employees working for firms offering data assurance

3 Turnover of firms engaged in data assurance activities active in the market

These metrics provide a sense of the magnitude of commercial activity in this space. The overall size of the market is an interesting metric in itself and also helps to build our understanding regrading potential gaps in provision.

There are certain caveats we need to bear in mind when interpreting the market sizing results:

! Our approach relies on firms' descriptions of their activity primarily via their own websites. This will capture the vast majority of firms offering relevant services but like any source of information it will not be 100% comprehensive. For example some firms may give very little detail of their service offering on their website and may therefore not appear in our list.

! Our study did not involve an empirical market definition exercise using competition economics. However, our chosen definition does have implications for our estimates of data assurance market size. We are focusing purely on suppliers of data assurance services and are also primarily interested in those firms who offer tailored support for clients / customers rather than just a technical platform.

91% of the 890 identified firms had employment information available. They vary significantly in size based on their employee base

1 In total, we identified 890 firms engaged in data assurance activities (for simplicity we refer to these as "data assurance firms" throughout, even though in some cases their commercial activity will not be entirely data assurance related

2 We have conservatively estimated that approximately 30,000 staff are involved in the delivery of data assurance activities.

Majority of identified firms have a relatively small number of employees each (see chart on next slide)	Data presented below is likely to be highly representative of all identified firms but caution is needed when interpreting it
▪ 56% of the firms who are engaged in data assurance activities had < 10 employees and 87% have < 50 employees. ▪ We identified seven firms who each employ > 10,000 staff (these are primarily large professional services firms). The majority of these employees in the largest firms will not be directly involved in data assurance activities.	▪ The information refers to headcount rather than full time equivalents. ▪ Data assurance activities are likely to be a subset of a wider package of services especially for the larger firms. However, it is not possible to separately estimate the proportion of their employees who carry out these activities. This in some cases may overstate the scale of data assurance activities within firms.

Majority of identified firms have a relatively small number of employees each

Breakdown of data assurance firms by employee size band
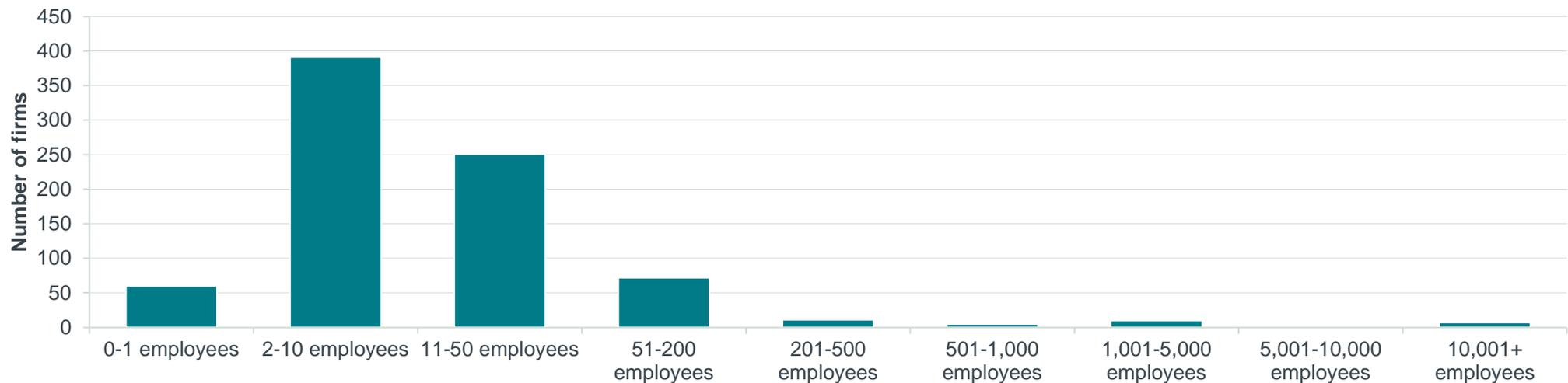
Source: Frontier analysis of data collected by glass.ai

Note: The total employment figure is based on the number of employees who have listed themselves as an employee of a firm on LinkedIn. We have assumed that only 1% of employees in firms with >1000 employees are engaged in data assurance activities. The firm level employment band data is a separate variable. These two separate measures of employment will be highly correlated but will not always be completely consistent.

# The size distribution of data assurance firms is skewed towards larger firms relative to the UK as a whole

Our sample of firms engaged in data assurance activities contain an above average rate of larger firms…	… and there are a number of reasons for this
▪ Recent ONS data on all tax registered UK firms suggests that 90% of all firms employ 9 staff or fewer, 8.5% employ 10-49 staff and the remaining 1.9% employ over 50 staff. ▪ Using these same categories of staff numbers, we find that in our sample of firms, larger firms are more common compared to the distribution of all UK firms. E.g. 31% of the firms we identified contain 10-49 employees as opposed to 8.5% of all UK firms.	▪ It is possible that a larger proportion of the smallest firms in the UK do not have an online presence and hence won't be picked up in our sample ▪ It may be that data audit type services tend to be offered by relatively large firms which is leading to the results we have highlighted.

Comparison of data assurance firms to all UK firms by employee size band

Legend (All UK Firms): ■ 0-9 ■ 10-49 ■ 50+

Legend (Data assurance firms): ■ 0-10 ■ 10-50 ■ 51+
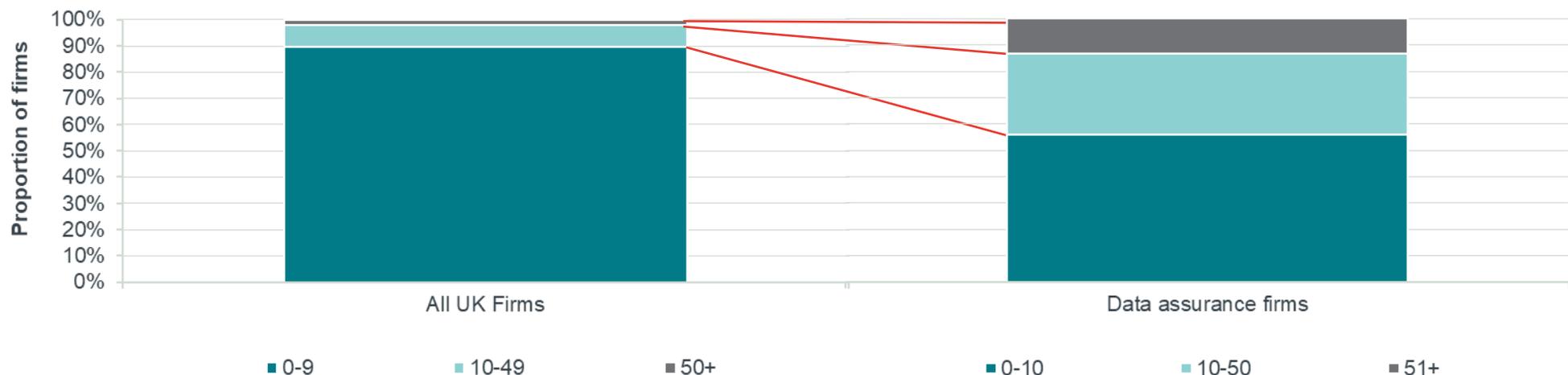
Source: Frontier analysis of data collected by glass.ai

Turnover is an alternative measure of size however our insights are not representative due to high levels of missing data on revenue

(3) **Turnover of firms engaged in data assurance activities active in the market**

| Data assurance firms who do report a turnover tend to be very large on average (see chart on next slide) |
|---|
| ■ Annual turnover data (sourced from publicly available Companies House information) is only available for a small subset of the identified firms. Specifically we have this metric for only 43 of our 890 firms (5%).

■ We have illustrated the proportion of firms engaged in data assurance activities (who published a turnover figure) in different size categories. We found that almost 80% of these firms have an annual turnover in excess of £5 million. |

Therefore turnover breakdowns will not be representative of the wider market
■ This low coverage rate is primarily because small companies are exempted from providing this information as part of annual accounts and our sample of firms are predominantly SMEs. ■ Large firms will offer a suite of services and data assurance services are only a subset. Therefore, some proportion of revenue will be generated as a result of non-data assurance activities. It is not possible to separately estimate the proportion of revenue that is linked to data assurance activities. Therefore we have reported total revenue.

Data assurance firms who do report a turnover tend to be very large on average

Breakdown of data assurance firms by turnover
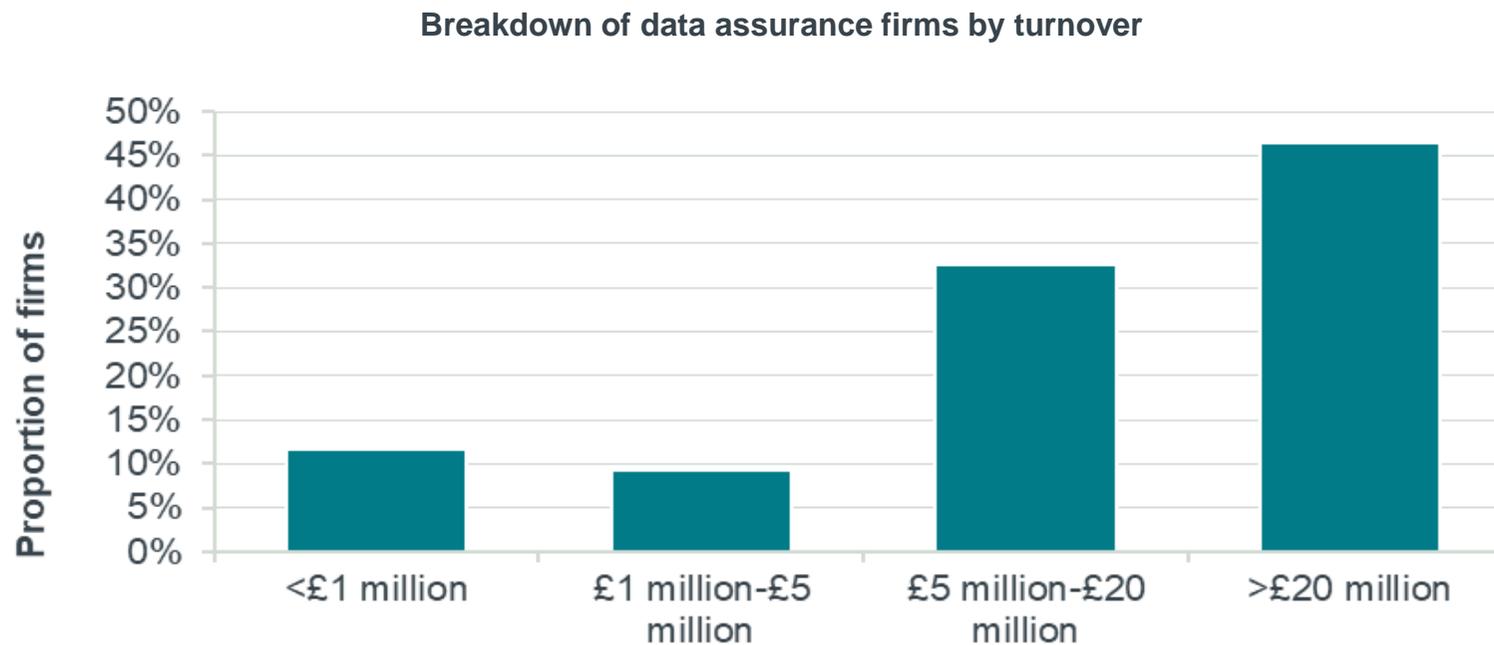
Source: Frontier analysis of data collected by glass.ai

Note: missing data mean that this chart is not representative of the entire data assurance market

Examining the rate of new entry into the data assurance market can allow us to asses market maturity

MATURITY OF MARKET

Our second research question focuses on the maturity of the data assurance market. We are interested in the extent to which data assurance service offerings are well established or rapidly emerging. The incorporation dates of firms that we have identified allow us to proxy market maturity.

Exploring the incorporation dates of data assurance firms allows us to identify trends in market growth

- We can examine when our data assurance firms incorporated in the UK for the first time. Therefore we can determine which of our final list of market participants "entered" the market in each year.
- We have sourced incorporation dates from publicly available Companies House records. Date of incorporation data is available for 588 of our 890 data assurance firms (66%).

There are certain caveats we need to bear in mind when interpreting the results:

! It is important to note that this rate of firm entry does not represent overall growth in the number of relevant companies. This is because we cannot capture firms which may have been present in earlier years but have subsequently exited the market and were therefore not identified via our search term methodology.

! In addition, in some cases larger firms may have initially incorporated a relatively long time ago to offer services which are not in scope of our work (e.g. accounting services) and started to offer data assurance services more recently. We are not able to determine this and therefore our estimates of market entry will not always correspond to actual commencement of data assurance offerings.

We found that 58% of the data assurance firms that we identified were incorporated in the last ten years

## Recent years have witnessed significant market entry (see chart on next slide)

- According to the latest ONS data (2020 B) the average birth rate of firms (new entrants as a proportion of active firms) was 12.7% per annum (between 2018-19). In relation to our data assurance firms specifically between 2014-2020 an average of 44 new firms entered the data assurance market each year (average birth rate of 11.7%).

- Therefore the entry of data assurance firms is roughly in line with that of the wider economy. This could potentially signal an opportunity for more rapid growth in the future if demand grows in line with trends in data availability. In total 58% of the data assurance firms that we identified were incorporated in the last ten years and 90% were incorporated after 2000.

This rate of firm entry does not represent overall growth in the number of relevant companies

- We cannot identify firms who have incorporated but subsequently left the market in previous years. However, our work does provide a useful benchmark in this regard and subsequent analysis can track this group of firms that we have identified longitudinally.

- We have also reported the equivalent figure of market entry for 2021 (10 firms) even though it is clearly not complete and therefore not fully comparable with previous years.

Recent years have witnessed significant market entry to provide data assurance services

Market entry over time
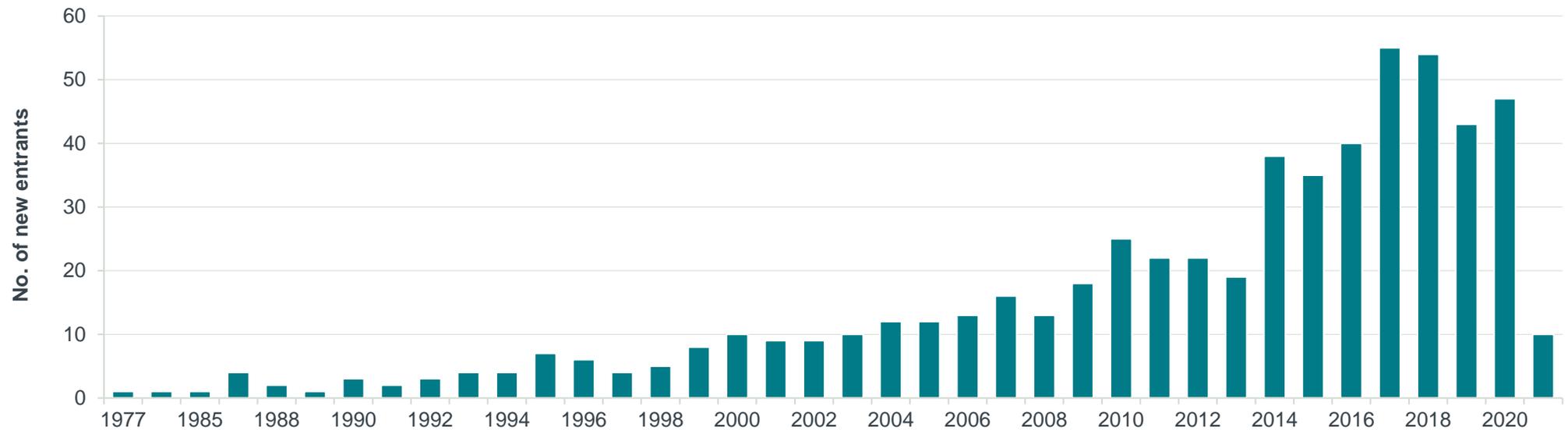
Source: Frontier analysis of data collected by glass.ai

Newer entrants are smaller than more established firms

The majority of new firms offering data assurance services are smaller businesses

▪ We have compared the more recent market entrants to all data assurance firms in terms of number of employees. Firms who have incorporated since 2018 are unsurprisingly small on average than more established firms. Specifically, 95% of the recent entrants employ 10 or fewer staff while the equivalent figure for all data assurance firms is only 56%.

▪ This does not imply that the recent market entrants play an unimportant role in the market. It is these firms which may be bringing innovative offerings to market to address previously unmet needs. In addition, more recent entrants may offer the highest potential for future growth and expansion in future years. Answering these two questions could be an important part of future research in this context.

Comparison of recent data assurance entrants (since 2018) to all data assurance firms by employee size band
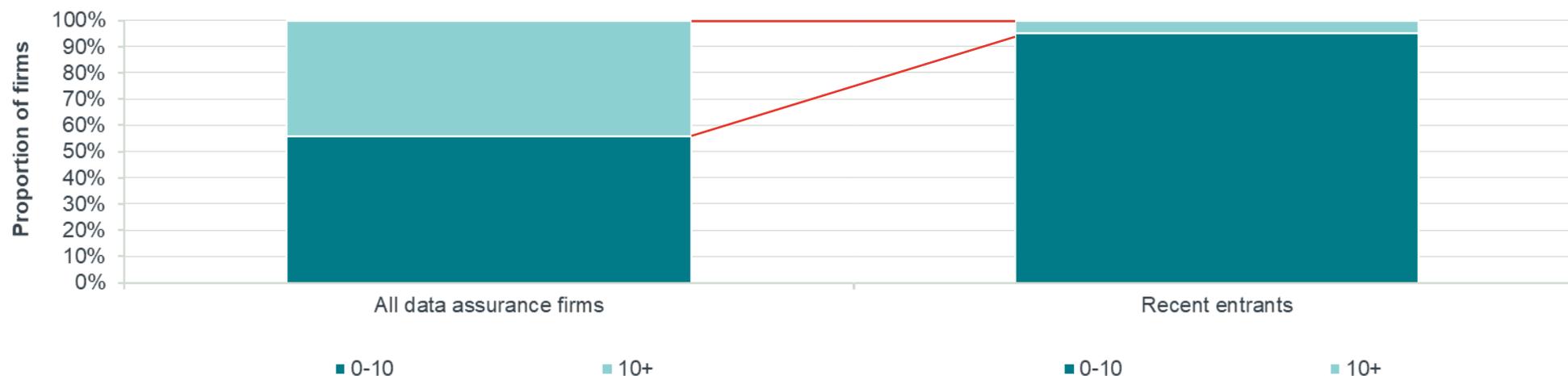
Source: Frontier analysis of data collected by glass.ai

Our third research question focuses on the sectoral and distribution of activities in the market (1/2)

TYPES OF FIRMS

Our third research question explores the sectoral distribution of data assurance firms.

Identifying the distribution of data assurance firms will help us to understand the type of economic activity that such firms are engaged in.

Sectoral distribution

- We have defined sectors according to **glass.ai's own classification system** (see Annex). This features 14 sector groupings (e.g. technology) and 109 sectors (e.g. computer hardware). This bespoke classification system is based on data read from the web - predicted from the company description and other web content.

- We have also defined sectors by the **Office for National Statistics' (ONS) Standard Industrial Classification (SIC) codes** The SIC is divided into 21 sections (e.g. manufacturing), which are then further broken down into 88 divisions (e.g. manufacture of textiles), 272 groups (e.g. manufacture of other textiles) and 615 classes (e.g. manufacture of carpets and rugs). SIC codes are used by ONS to classify businesses according to the type of economic activity in which they are engaged. The two classification systems are not directly comparable

There are certain caveats we need to bear in mind when interpreting the results:

Data assurance firms will not all be included under a single ONS SIC code class or glass.ai sector. The **different classifications of the data assurance firms further demonstrates the difficulty in isolating relevant services within a single traditional market.** This is because data assurance activity can play a role across many traditional sectors and stages of the value chain.

Note: We have used both of these classification systems because they offer distinct advantages. glass.ai's classification is more recent and therefore may represent a more up-to-date breakdown of firms. However, SIC codes are widely used as part of official statistics which allows us to compare data assurance firms to the rest of the economy.. Both of these distributions indicate the range of sectors that data assurance firms are active in.

Our third research question focuses on the sectoral and distribution of activities in the market (2/2)

TYPES OF LOCATIONS

Our third research question also explores the geographical distribution of data assurance firms.

Identifying the distribution of data assurance firms will help us to understand the locations of economic activity that such firms are based in.

Geographical distribution

This will **provide an indication of which parts of the UK are flooded with data assurance firms and which areas have less provision**. However, many of these types of services may not require physical interaction so potential customers across the UK may be able to tap in the offerings of firms who are not necessarily physically proximate

There are certain caveats we need to bear in mind when interpreting the results:

Some data assurance firms might have multiple offices across UK. We have identified their main offices based on data read from the web - predicted from the company description and other web content.

Data assurance activity is spread across many different clusters of economic activity as we would expect

**Data assurance firms are most commonly found in "technology sectors"**

▪ A glassi.ai sector grouping and sector is available for all of our 890 data assurance firms. **Using this glass.ai classification system we have found that 67% of data assurance firms identified in our sample are in the Technology sector grouping**. The top three sector groupings (this is the more aggregated level of economic activity) Technology, Professional Services, and Financial Services collectively account for 96% of the firms in our sample firms (we have also explored the more granular sector breakdown which is presented on the next page).

▪ SIC codes are available for 579 of our data assurance firms (65%). **The most commonly appearing SIC code divisions (two digit code) was Computer programming, consultancy and related activities**. The top three most frequent two-digit SIC codes in our data (Computer programming, consultancy and related activities, Activities of head offices; management consultancy activities and Office administrative, office support and other business support activities) accounted for 77% of the data assurance firms.

Top three glass.ai Sector Grouping: percentages
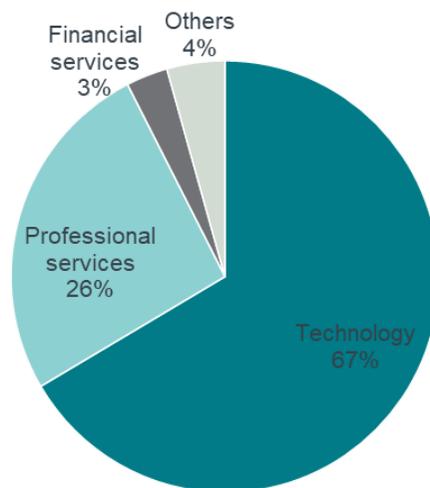
Source: Frontier analysis of data collected by glass.ai

Top three two digit SIC codes: percentages
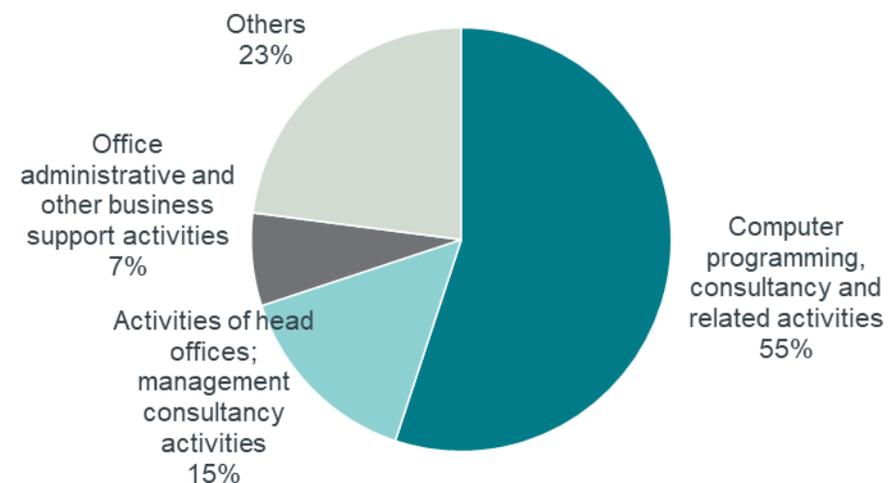
Source: Frontier analysis of SIC data by ONS

We cannot isolate data assurance activities within a single traditional sector (1)

Data assurance firms specialise in technology, finance and professional service areas

- Previously showed a high level classification of data assurance firms' sectoral composition. We have also examined a more granular breakdown to better understand the specific types of economic activities that the data assurance firms in our sample are undertaking.

- Using the glass.ai classification system we found **the most common sectors, included Information Technology and Services, Consulting, and Computer Software.** This illustrates the wide range of sectors that data assurance companies are active in. It also reinforces the point that we cannot isolate data assurance activities within a single traditional sector.

- We have also examined the sectoral mix of recent market entrants (firms that incorporated since 2018). Overall the breakdown is relatively similar to the pattern we observe amongst all data assurance firms. **Recent entrants are slightly more likely than all data assurance firms to be classified in the Information Technology and Services sector (48% vs 44%) and slightly less likely to be in the Accounting sector (1% vs. 2%).**

Top ten glass.ai sub-sectors: percentages

Information Technology and Services, 44%
Consulting, 14%
Computer Software, 13%
Computer Networking and Security, 8%
Others, 5%
Information Services, 4%
Security and Investigations, 3%
General Financial Services, 3%
Accounting, 2%
Law Practi... and Servi...
Oil and Ene... 1%
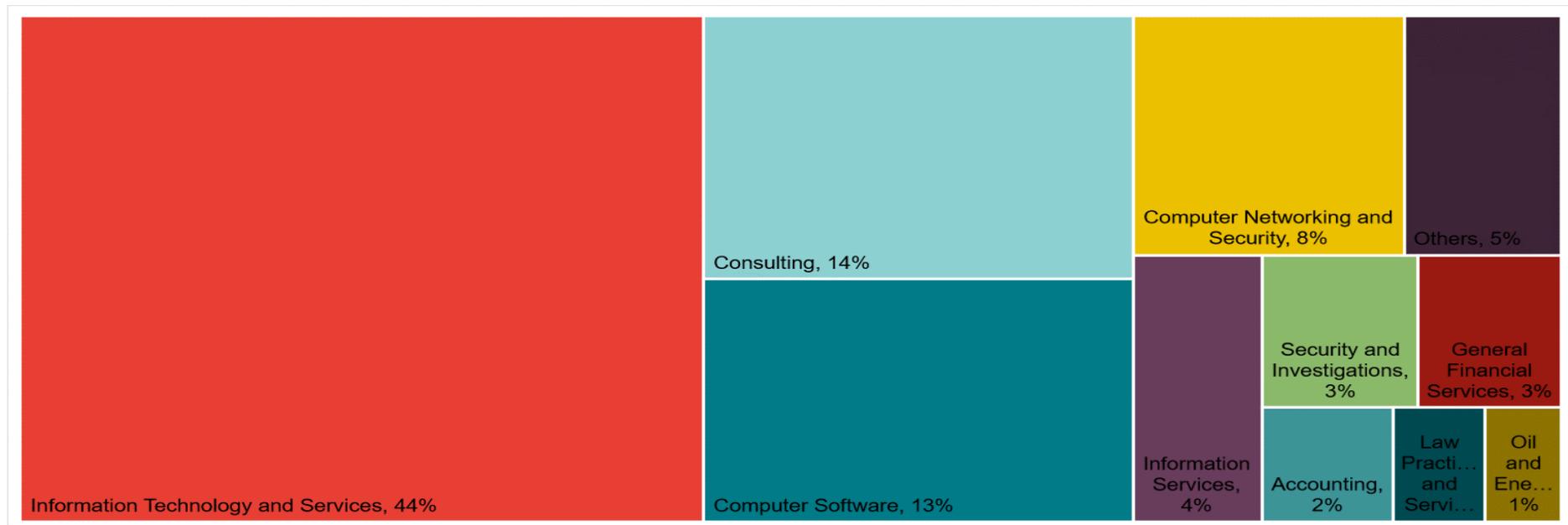
Source: Frontier analysis of data collected by glass.ai

We cannot isolate data assurance activities within a single traditional sector (2)

A similar pattern can be seen using granular SIC classifications

- We have examined the most common five digit SIC classifications amongst our data assurance firms. Collectively the top 10 5 digit SIC codes account for 86% of firms identified. Computer consultancy services, management consultancy services, data processing activities and business support services all feature prominently.

Top ten SIC sub-sectors (5 digit classes): percentages

Other IT and computer service activities, 15%

Management consultancy activities, 14%

Business and domestic software development, 11%

Other professi... scientific and technical, 3%

Data processing, hosting, 2%

Other information service activities, 2%

Computer consultancy activities, 28%

All others, 14%

Other business support service activities, 7%

Other software publishing, 2%

Activities of extraterrit... organisati... 2%

Source: Frontier analysis of SIC data by ONS

Data assurance firms are over-represented in sectors such as computer programming, and management consulting

We can see that the sectoral breakdown of data assurance firms differs from the overall distribution of all UK firms

- We have used the overall breakdown of UK firms as a benchmark for the classification of data assurance firms. We can see that the sectoral breakdown of data assurance firms differs significantly from the overall distribution of all UK firms.

- In the table below we compare the percentage of data assurance firms appearing in the six SIC code divisions that contain the highest number of these firms with the percentage of all UK companies in these divisions. **We see that data assurance firms are over-represented in sectors such as computer programming, management consulting and information service activities relative to all UK companies**.

SIC division	Data assurance companies	All UK companies (ONS data)
Computer programming; consultancy and related activities	55%	6%
Activities of head offices; management consultancy activities	15%	7%
Office administrative and other business support activities	7%	6%
Information service activities	5%	0.4%
Other professional, scientific and technical activities	4%	3%
Publishing activities	2%	0.5%

Source: Frontier analysis of SIC data by ONS

Firms classified in the accounting sector tend to be larger than other data assurance firms

Accounting firms tend to have a higher number of employees on average

- We have also examined the interactions between firm size and sectoral classification. Using the glass.ai classification system we can see that accounting firms are overrepresented in the larger size bands (all categories which cover more than 50 employees). It is these audit type firms who are therefore most likely to be providing data assurance services alongside other service offerings. These firms also tend to be older on average.

Sub-sectoral distribution of firms by number of employees

Legend:
- Information Technology and Services
- Consulting
- Computer Software
- Computer Networking and Security
- Information Services
- Security and Investigations
- General Financial Services
- Accounting
- Law Practice and Services
- Oil and Energy
- Others
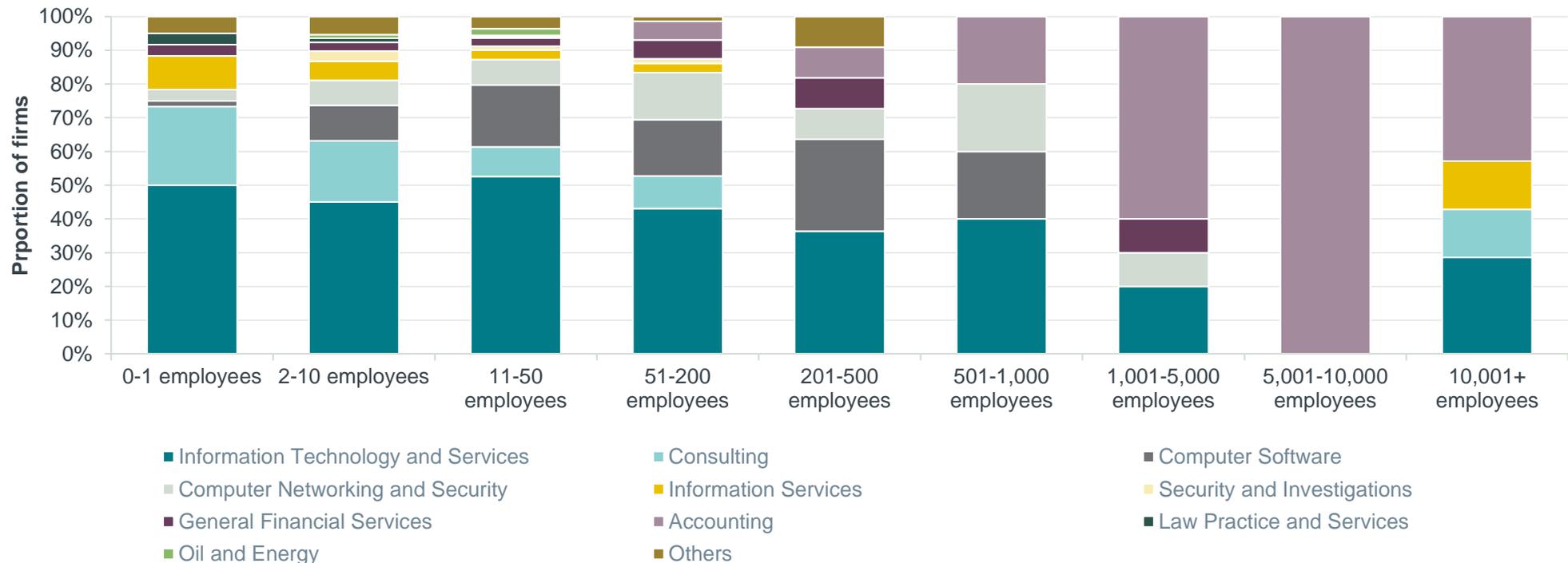
Source: Frontier analysis of data collected by glass.ai

Certain regions such as London contain a disproportionate share of data assurance companies

There is a clear clustering of firms around London which contains 40% of the firms in our sample

- We have also examined where the data assurance firms identified in our sample are located in the UK using available location indicators drawn from Companies House and LinkedIn. Overall, 69% of firms in our sample have data which we can use to determine the relevant UK regions. From the table below, we see that there is a clear clustering of firms around London which contains 40% of the firms in our sample. This is more than twice the proportion of all UK firms who are based in London.

- **This may suggest the emergence of a data assurance cluster in London. This is in keeping with wider tech industry location trends**. Previous research by Tech Nation (2021) found that London was fourth for tech VC investment globally in 2020 and is for example the second biggest Fintech hub. The data we have used may exaggerate this somewhat if large firms with multiple sites are more likely to report their London address than other locations

Region	Data assurance providers	All UK companies
London	40%	19%
South East	17%	15%
North West	7%	10%
West Midlands	7%	8%
East of England	6%	10%
South West	6%	9%
Scotland	5%	7%
East Midlands	3%	7%
Yorkshire and The Humber	3%	7%
North East	3%	3%
Wales	2%	4%
Northern Ireland	1%	3%

Source: Frontier analysis of data collected by glass.ai

frontier economics

Data assurance companies are located in urban areas across the UK

Overall, we have been able to gather location data for 70% of the firms in our sample.

- We have mapped the precise location of each data assurance firm using the postcodes associated with the address provided in their Linkedin profiles/websites.

- In line with the regional results we present on the previous page we see that there is a clear grouping of firms around London and the south of the country.

- There are also pockets of data assurance activity in all regions which tend to be clustered around urban centres.

Geographic distribution of UK Data assurance firms

Region
- East Midlands
- East of England
- Isle of Man
- London
- North East
- North West
- Northern Ireland
- Scotland
- South East
- South West
- Wales
- West Midlands
- Yorkshire and The Humber
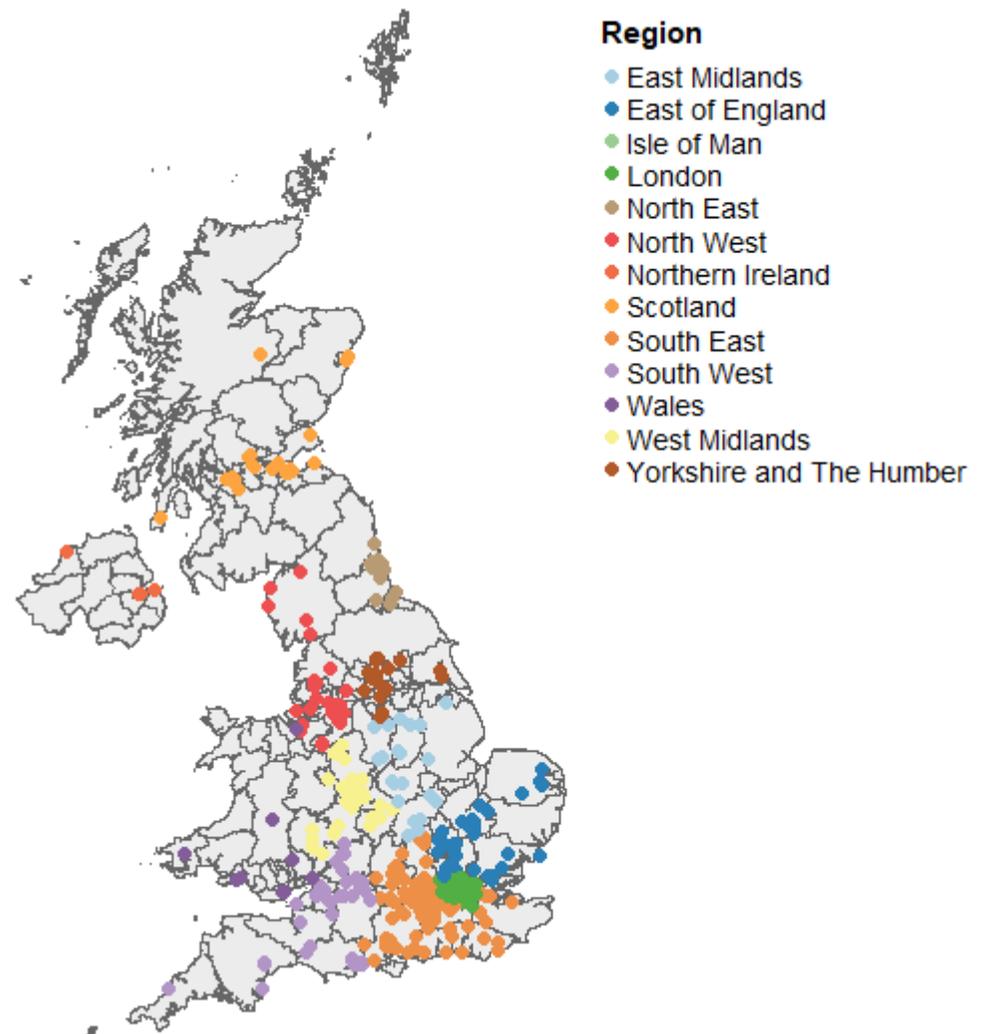
Source: Frontier analysis of data collected by glass.ai

Examining the number of firms that matched against search terms allowed us to highlight common assurance services

TYPES OF PRODUCTS AND SERVICES

Our last research question focusses on the success of search terms in identifying firms as part of our web crawl exercise.

We are interested in the type of products and services that are offered by the firms active in the data assurance market and the topic matches can provide an overall indication of this.

Exploring which firms include specific search terms in their descriptions of their offerings provides an indication of data assurance offerings present in the market.

- We defined a list of keywords we then used to identify data assurance firms. We based this list on a rapid desk research exercise, our own experience and conversations with ODI. It represents a robust and coherent list of potential activities, products and services but may not be entirely comprehensive given the wide range of terminology used in this context.

- In addition, we also carried out specific deep dives into larger 'accounting' firms who were knew to be active in this market. This allows us to better understand specific business moils and service offerings that are in place currently.

**There are certain caveats we need to bear in mind when interpreting the results:**

It also may be the case that firms do offer categories of products/services which are not explicitly included in their website. If this is the case they will not be picked up by our analysis.

The majority of firms we identified may focus on a specific niche within the market

Overall, most firms may have narrow data assurance offerings

▪ As part of the strategy to collect data on data assurance firms, we had identified 75 keywords / phrases. All of the data assurance firms we identified matched against at least one of these 75 search terms.

▪ 77% of the firms we identified only matched against a single keyword / phrase, 16% matched against two terms and the remaining 7% matched against three or more terms. **This provides tentative evidence that the firms we have identified tend on the whole to focus on a specific niche within the topic of data assurance primarily (e.g. verification of source data or prevention of data leakage), rather than trying to provide a more all encompassing offer.** Further research will be needed to confirm this

▪ We have also examined the number of topic matches of recent market entrants (firms that incorporated since 2018). **Data assurance firms who have incorporated more recently are slightly more likely to only match against a single of our keywords than all data assurance firms (80% vs. 77%).** This may provide tentative evidence that newer firms are slightly more likely to be focusing on a specific niche of data assurance relative to more established firms.

Proportion of firms matching with different number of topics
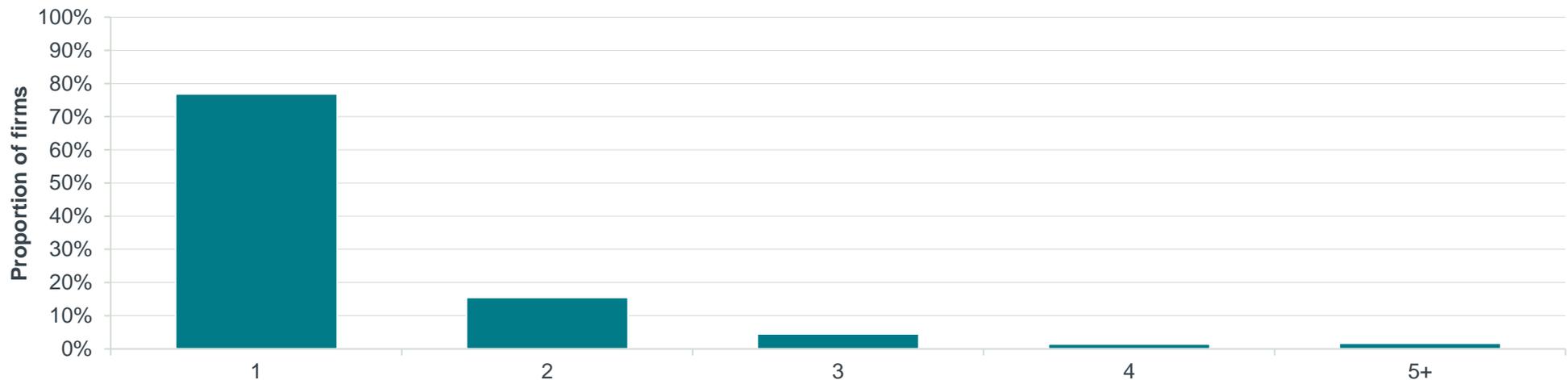
*Source: Frontier analysis of data collected by glass.ai*

52 out of 72 search terms identified all the firms in our sample

Successful search terms can help to identify the types of products and services offered in the market

Most of the data assurance firms are identified through a subset of topic matches

- **72% of the firms we identified were linked to one of the top 10 most frequent topics**. 'Data management', 'Data governance' and 'Data Security' were the most frequently matched topics. The specific term 'data assurance' matched with only 8 firms. These firms also matched against other terms- 'data management', 'data governance', 'data security', IT governance', 'data quality' and 'data breach'.

- **The top 10 list included search terms from all five our search term categories which suggests the identified firms use a broad vocabulary to describe their offerings.** Unsurprisingly it is the search terms from the "overarching category" which identified the most firms. These more generic terms will capture more organisations than specific articulations of products or services. Some of these terms will be synonymous with each other (data management and information management will likely refer to very similar service offerings). In other cases data assurance firms may be describing the same problem from two different perspectives. For example, two firms may offer very similar services but one emphasises how they can increase quality while another focuses on the avoidance of risk.

- We have also examined the frequency of successful matches amongst recent market entrants (firms that incorporated since 2018). The top-10 list is broadly similar. However, **data compliance and data integrity both feature in the top 10 of most commonly appearing terms when we look at recent market entrants only and do not feature in the overall list above**. This could represent a slight shift in the type of offerings amongst newer firms.

Data management and data governance were the two most commonly identified search terms

Search categories
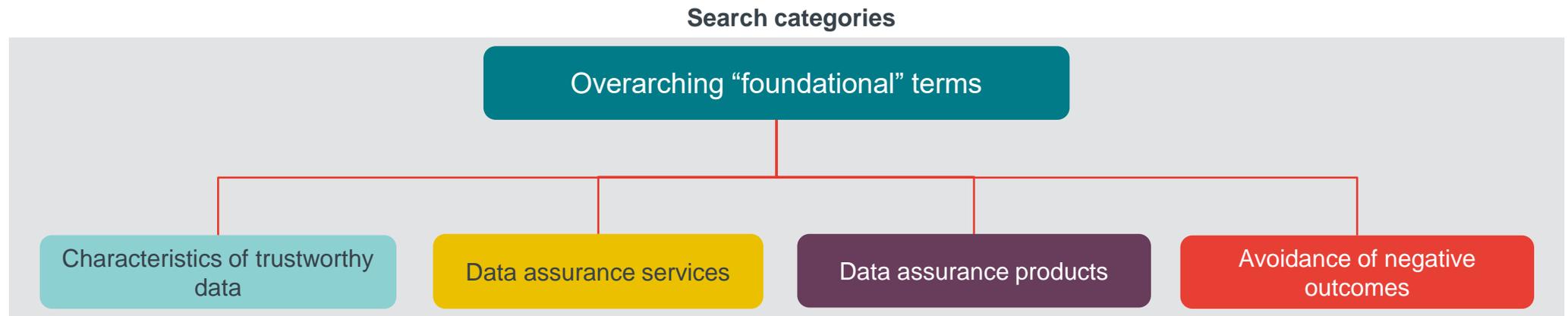
Proportion of firms identified against top 10 search terms

Search term	Number of firms identified	% of total
Data management	168	13%
Data governance	124	10%
Data security	122	10%
Information management	95	8%
Data quality	92	7%
Information assurance	85	7%
Data strategy	68	5%
GDPR compliance	64	5%
Data sharing	51	4%
Master data management	43	3%

Terms such as data accreditation and data validation did not match against any firms

Unsuccessful search terms can help highlight potential gaps in the market for data assurance

Overall, 23 search terms did not identify any firm

- Unsuccessful search terms included:
  - **Foundation terms** such as data accreditation and data validity
  - **Characteristics of trustworthy data** such as data guarantee and data confidence
  - **Services** such as computer assisted auditing techniques
  - **Products** such as data documentation
  - **Avoidance of negative outcomes** such as data loss prevention techniques

- This implies that these type of offerings are not explicitly referred to by data assurance firms themselves. This could provide tentative evidence of gaps in the market and could also reflect differences in vocabulary used across the industry. Likewise some of the terms which only matched against a small number of firms could also be potential gaps that are worthy of further exploration for ODI.

Our list of data assurance firms includes multiple examples of offerings at different stages of the value chain

We can use the data value chain to highlight examples of firms' data assurance offerings

As described previously the data value chain describes various value adding process from initial collection of data through to its use and exploitation. Data assurance offerings exist currently in the market at each stage of this conceptual value chain. We have provided examples below which are taken directly from firms' own descriptions and are illustrative rather than exhaustive.

Data value chain

| Data collection & sourcing | Data harmonisation | Data sharing/ publishing | Data interlinking | Data exploitation | Data curation |
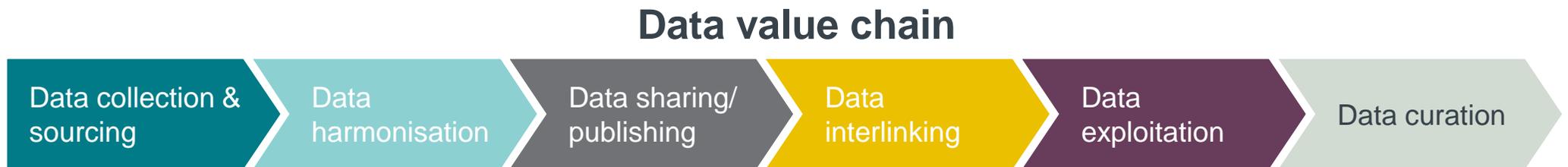
Source: adapted from Attard et al. (2016) and Attard et al. (2017)
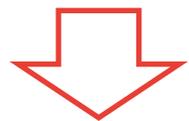
- "We provide resources, skills and technology to access better data faster. We assist by delivering an integrated data solution from forming a data strategy to the automation of **data collection**, data cleansing, data classification"
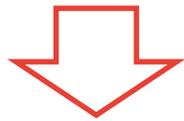
- "Data consultancy with a mission to empower you and your data. Services include data set-up and **harmonisation**"
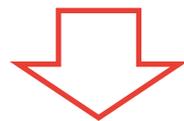
- "Empowers organizations to unlock more value from their data and to accelerate data-driven business models through its secure, custom-branded data **sharing and collaboration platform** (delivered as a service."
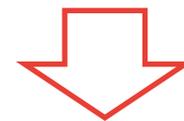
- ""We allow for collaboration across organisational and technological boundaries. Our products and services use semantic web techniques and **linked** data principles to improve the search and discovery of information."
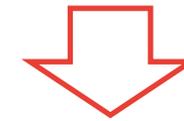
- "Data quality solutions that improve customer engagement and help you **make smarter decisions** using the most accurate data"
- "Our mission is to help **clients derive value** from data and make data driven decisions"
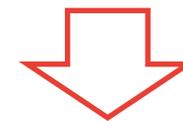
- "We tailor our data security systems and solutions to meet your businesses every requirement. The solutions ensure **prevention of data leakage**, whether at rest within the network, mobile device or storage or indeed data in transit across public or private networks."

We carried out a deep-dive crawl of professional services firms

Professional services firms will compose an important segment within the data assurance market

- Before carrying out the web crawl we knew that medium to large professional services firms offer some data assurance services (in addition to traditional financial auditing services or broader technology based consulting). We therefore carried out a deeper crawl across known UK professional services firms to:

  o establish which of these firms offered services which were in scope of our review; and

  o highlight the specific products and services that they offer and the business models they have adopted.

- This deep dive crawl consisted of three steps. This provided more granular and detailed data on the specific data assurance offerings of these 15 firms than we have for the remainder of the sample.

Figure: Steps in the deep crawl exercise

1 Identify all potential professional services firms

Firstly over 100 professional firms were identified

2 Identify which of these firms offers relevant services

Our deep crawl then found that 15 of these firms offered relevant services according to their website

3 Collect multiple pieces of information in terms of service offerings from each firm

The deep crawl then collected multiple pieces of evidence from the websites of these 15 companies. The top ranked evidence (up to four pieces per organisation has been included in our final dataset)

Services offered by professional services firms tended to fall into a number of categories

We have grouped the service offerings regarding data assurance activities described by professional services firms into a number of groups:

Process focused	Outcome focused

Category of offering / capability

Improving data governance	Data privacy & security	Data compliance	Increasing data quality	Sharing data	Making best use of data

Example description

- Our clients have: "*challenges with respect to (further) setting up data governance, master data management, data quality management and data lineage across the entire organization*"

- Our "*services cover a wide range of risk scenarios such as employee error, algorithmic bias, faulty algorithm design, malicious manipulation and poor data quality, all of which can lead to non-compliance with data protection and privacy regulations*"

- "*Projects may be undertaken to provide assurance for the Board and other stakeholders on data quality or entity status.*"

- "*Improving data accuracy and creating valuable insights for our client*"

- *Data quality solutions continue to mature through the addition of more augmented capabilities than before, driven by growing needs in existing and new use cases*

- "*Data management across data ecosystems require significant time and investment just to achieve consistency and traceability to improve trust with customers*"

- "*We are able to help companies manage data, identify and mitigate issues such as bias or unnecessary access that could compromise data integrity, and drive increased confidence in insights driven by AI* "

The professional services firms we identified refer to different combinations of data assurance offerings

Professional services firms will compose an important segment within the data assurance market

- In the table below we have illustrated which of the 15 professional services firms included content on their website that referred to one of the six topic categories which we defined.
- We can see that different firms use different terminology to describe their services and occupy different places in the market. The most common category of capability / offering relates to data governance.

Accountancy firms identified	Improving data governance	Data privacy & security	Data compliance	Increasing data quality	Sharing data	Making best use of data
EY	✓	✓	✓		✓	
KPMG UK	✓			✓		✓
Gartner	✓			✓		
RSM UK	✓		✓			✓
Menzies LLP		✓				
Mitchell Charlesworth		✓				
Russell Bedford International				✓		
BDO UK LLP	✓					
DNS Accountants	✓					
Praxity Alliance	✓					
Crowe UK	✓					
Grant Thornton UK LLP	✓		✓			✓
Deloitte	✓					
Mazars in UK	✓					
PwC UK	✓			✓		✓

Source: Frontier analysis of data collected by glass.ai

We have also examined the most common business models used by data assurance firms

We can provide a high level indication of the business models that are being used by data assurance firms

- To explore the types of commercial model employed by data assurance firms we have examined the content of each data assurance firm's website home page. This will provide an indication of how data assurance firms earn revenue and what type of organisations they typically offer their services to.

- We firstly defined a list of 17 business models based on our previous experience and existing categorisations.1 We then crawled the website homepages for references to these known business models. This allowed us to determine the frequency with which certain commercial operating models are being used.

- Firms may operate a mixed model with multiple revenue streams. Therefore it is possible that firms match against multiple business models categories. In total we business model signals were picked up for 710 companies (80% of our list).

List of businesses model categories
Advertising
Affiliate marketing
Bundling
Consulting
Digital goods
Direct sales
Distribution
E-commerce
Franchise
Leasing
Licensing
Manufacturer
Marketplace
Peer-to-peer
Product to service
Retailer
Subscription

1 See for example: https://www.nerdwallet.com/article/small-business/what-is-a-business-model

The majority of data assurance firms are consultancies. Firms also note targeted market sectors and revenue generation routes (1/2)

Consulting is by far the most common business model identified (669 firms). It involves building up experience and expertise in specific domains of data assurance and then carrying out work for clients.

E.g. "*We are dedicated to protecting client's reputation by providing strategic and pragmatic advice to provide information security*"

1 **Customer base**

Many data assurance companies talk about the clients they serve (who themselves will adopt a variety of business models).

Examples include:

- **retailers (94 data assurance firms). E.g.** "We provide the full suite of data management services (advisory, consultancy, Master Data Management (MDM), Product Information Management (PIM), Data Quality. Client examples include leading Retailers, Manufacturers, Government Agencies"

- **distributors (73 data assurance firms). E.g.** "We are a data security, governance, regulation and compliance (GRC) solution provider. Our systems help to protect over a million users in areas as diverse as security services, manufacturing, and distribution"

- **advertising platforms (80 data assurance firms). E.g.** "specialise in design and delivery of information management software and technology products and platforms. We provide consultancy services to the Information Technology, Advertising, Marketing, Design and Media and Transport industries"

Other examples of customer segments include e-commerce, finance and healthcare

The majority of data assurance firms are consultancies. Firms also note targeted market sectors and revenue generation routes (2/2)

(2) Revenue generation

Other firms offer a subscription or licencing service to clients which features a mix of platform access and ongoing sharing of expertise.

This includes:

- **Subscription services (72 data assurance firms). E.g.** "We cover Compliance, Security, Information Governance and Contracts. Our services are accessible and cost effective with a monthly subscription service that allow you to choose the packages and package components that apply to you"

- **Licencing services (37 data assurance firms). E.g.** "The company behind the new standard for data security. We commercially exploit our IP through licensing agreements, and by offering a base product and implementation design book. Our team also brings the opportunity for consultancy, development and industry thought-leadership"

Our work has provided a comprehensive overview of the firms providing data assurance services in the UK and their characteristics

1

Our work provides useful insights into the market for data assurance

Frontier Economics and glass.ai were commissioned by ODI to carry out a detailed study of the size, features and characteristics of the market for data assurance services in the UK.

The novel dataset we have created and the subsequent analysis we have carried out has highlighted a number of interesting findings:

- We identified 890 providers who explicitly refer to data assurance capabilities or offerings. This market has been growing steadily. Recent entrants may offer opportunities for growth in the coming years.

- Data assurance firms offer a variety of services to customers in multiple markets. Some large accounting firms do offer a range of data assurance services alongside more traditional financial audit offerings.

- Data assurance firms are disproportionately based in London. Other parts of the country may be less well served.

- Some types of offering such as data accreditation for example are not explicitly referred to by data assurance firms themselves. This could provide tentative evidence of gaps in the market and could also reflect differences in vocabulary used across the industry.

Our work also provides a direction on where the data assurance market can benefit from further research

2

Further research could build on the work we have carried out

Some research questions were beyond the scope of our study. We did not engage directly with data assurance providers/their customers who could provide insights on:

Some interesting research questions were beyond the scope of our study. We did not engage directly with data assurance providers or their customers who would be able to provide insights on:

- Demand for data assurance:
  - What are key user needs in this context?
  - What are current levels of customer satisfaction?
  - Are there consumption externalities?

- Supply of data assurance
  - What are the latest trends in provision?
  - What are the barriers to further expansion?

References (1)

- Attard, J., Orlandi, F. and Auer, S., 2016, October. Data value networks: enabling a new data ecosystem. In 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (pp. 453-456). IEEE.

- Attard, J., Orlandi, F. and Auer, S., 2017, March. Exploiting the value of data through data value networks. In *Proceedings of the 10th international conference on theory and practice of electronic governance* (pp. 475-484).

- Digital Competition Expert Panel (2019) Unlocking digital competition https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel

- Digital Curation Centre's (2010) Data Accreditation overview https://www.dcc.ac.uk/sites/default/files/documents/resource/briefing-papers/Abbott%20Data%20accreditation.pdf

- Frontier Economics (2020) Geospatial Data Market Study Report https://www.gov.uk/government/publications/enhancing-the-uks-geospatial-ecosystem/frontier-economics-geospatial-data-market-study-report-executive-summary

- Frontier Economics (2021) Economic Impact of Trust in Data Ecosystems https://theodi.org/article/the-economic-impact-of-trust-in-data-ecosystems-frontier-economics-for-the-odi-report/

- Gartner (2021) Magic Quadrant for Data Quality Solutions https://www.gartner.com/doc/reprints?id=1-25D4TRC9&ct=210302&st=sb

- ICO (2021 A) Certification https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/certification/

- ICO (2021 B) Data sharing code of practice https://ico.org.uk/media/for-organisations/data-sharing-a-code-of-practice-1-0.pdf

References (2)

- ICO (2021 C) Guide to the UK General Data Protection Regulation (UK GDPR) https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

- ISO (2021) ISO/IEC 27001: Information Security Management https://www.iso.org/isoiec-27001-information-security.html

- ODI (2021 A) R&D: Building trust through audit and certification https://theodi.org/project/building-trust-through-audit-and-certification/

- ODI (2021 B) What is assurance? https://theodi.org/article/assurance-trust-confidence-what-does-it-all-mean-for-data/

- ONS (2020 A) UK business: activity, size and location https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/datasets/ukbusinessactivitysizeandlocation

- ONS (2020 B) Business demography, UK: 2019 https://www.ons.gov.uk/businessindustryandtrade/business/activitysizeandlocation/bulletins/businessdemography/2019

- Oxford Insights (2019) Exploring Data Trust Certifications https://theodi.org/wp-content/uploads/2019/04/Report_-Exploring-Data-Trust-Certification.pdf

frontier economics

Search terms (1)

Overarching "foundational" terms

- Data assurance
- Information assurance
- Information management assurance
- Data governance
- IT governance
- Data certification
- Data accreditation
- Data validation
- Data validity
- Trust in data processes
- Data sharing
- Data ethics
- Digital trust
- Data control
- Data management
- Data assessment

This list includes high level terms that firms might use to describe an overarching suite of offerings in this space.

Multiple firms may describe very similar offerings in slightly different ways so we have included as many synonyms as possible.

Search terms (2)

Characteristics of trustworthy data

- Data accuracy
- Data standards
- Data confidence
- Data guarantee
- Data accountability
- Data standardisation
- Data quality
- Data reliability

This list includes features of trustworthy data that can be assessed / certified as part of a data assurance product or service

Search terms (3)

Data assurance services

- Data auditing
- Data strategy
- Data audit
- Data security
- Data curation
- Data protection audit
- Data quality assessment
- Computer Assisted Auditing Techniques
- Audit data analytics
- Augmented data quality
- Data profiling
- SOC 2
- ISO / IEC 27001
- Certified Information Security Manager
- Certified Information Systems Auditor / Professional

This list includes specific services which may be offered by data assurance firms

Search terms (4)

Data assurance products

- Metadata management

- Data documentation

- Master data management

- Data catalogue

- Reference data management

This list includes specific elements of data assurance platforms

Search terms (5)

Avoidance of negative outcomes

- Data compliance

- Data remediation

- Data loss prevention techniques

- Data breach

- GDPR compliance

- Data leakage

This list includes negative outcomes that data assurance process can help to avoid

The web crawling process can be further broken down into four sub-steps

Data exploration

- Agreement on keywords
- Algorithm reads 2 million UK business websites
- Discovery of additional sources

Intelligent crawling

- Gathering and structuring of open web data
- Reading social media and news sites to augment results

Analysis & QA

- Automatic / manual QA processes
- Removing anomalies
- Match with official data sources

Data delivery

- Packaging of data
- Sharing of final CSV

We implemented two additional criteria to improve precision without compromising on comprehensiveness

Rules to improve accuracy of quantitative data collection

▪ All of the search terms we defined could in certain circumstances be used by a business to refer to an activity or service offering that is in-scope for the purposes of our work. However, every firm which includes any of these key words on their website will not necessarily be relevant. This is because certain aspects of data assurance terminology may be used inconsistently across the landscape for a wide range of in-scope services/products as well as more generic or less relevant offerings. We therefore implemented two additional rules to increase precision and limit the amount of manual input and ad-hoc decision making required to reach a final dataset:

Exclude firms from peripheral sectors:

An initial manual review of a sample of data revealed that the vast majority of firms classified in certain sectors of the economy tended to be relevant for our work (e.g. Information Technology and Services, Computer Networking and Security, Consulting) whereas the bulk of firms in other sectors (e.g. Healthcare) tended to be "false" matches. We therefore primarily focused on a subset of sectors which yielded a high match rate during this initial review stage. Firms which matched against one of our search terms but were not classified in one of those core sectors were not taken forward into the final dataset.

Unless they match against multiple search terms:

We were conscious that there may be a small number of relevant firms in non-core sectors who we did not want to exclude. Therefore, all firms which matched against more than one search term (regardless of sector) was included in the final dataset. Firms which included more than one of our terms on their website are far more likely to be in scope for our work overall.

▪ These two rules offer a pragmatic balance between precision and comprehensiveness.

glass.ai firm sectoral classification (1)

Sector Grouping	Sector
Consumer Goods and Services	Apparel and Fashion
Consumer Goods and Services	Consumer Electronics
Consumer Goods and Services	Consumer Products and Services
Consumer Goods and Services	Cosmetics and Toiletries
Consumer Goods and Services	Food and Beverages
Consumer Goods and Services	Furniture
Consumer Goods and Services	Luxury Goods and Jewellery
Consumer Goods and Services	Retail
Consumer Goods and Services	Sporting Goods
Consumer Goods and Services	Textiles
Consumer Goods and Services	Tobacco
Consumer Goods and Services	Wholesale
Consumer Goods and Services	Wine and Spirits
Energy and Environmental	Environmental Services and Conservation
Energy and Environmental	Oil and Energy
Energy and Environmental	Renewables and Sustainability
Energy and Environmental	Utilities
Financial Services	Banking
Financial Services	General Financial Services
Financial Services	Insurance
Financial Services	Investment Banking and Advisory
Financial Services	Investment Management
Financial Services	Venture Capital and Private Equity
Government	Central and Local Government
Government	Government Agencies and Other Public Bodies
Government	Political
Healthcare and Scientific	Alternative Medicine
Healthcare and Scientific	Biotechnology
Healthcare and Scientific	Hospitals and Medical Practices
Healthcare and Scientific	Life Sciences and Pharmaceuticals
Healthcare and Scientific	Medical Devices
Healthcare and Scientific	R&D and Scientific
Healthcare and Scientific	Veterinary

Sector Grouping	Sector
Industrial and Agriculture	Agribusiness and Fishery
Industrial and Agriculture	Aviation, Aerospace and Defense
Industrial and Agriculture	Chemicals
Industrial and Agriculture	Electrical and Electronic Manufacturing
Industrial and Agriculture	Industrial Automation
Industrial and Agriculture	Machinery
Industrial and Agriculture	Mechanical and Industrial Engineering
Industrial and Agriculture	Mining and Metals
Industrial and Agriculture	Packaging and Print
Industrial and Agriculture	Paper and Forest Products
Industrial and Agriculture	Plastics
Leisure and Hospitality	Entertainment and Leisure
Leisure and Hospitality	Events Services
Leisure and Hospitality	Gambling and Casinos
Leisure and Hospitality	Health, Wellness and Fitness
Leisure and Hospitality	Hospitality and Restaurants
Leisure and Hospitality	Leisure, Travel and Tourism
Leisure and Hospitality	Recreational Facilities and Services
Leisure and Hospitality	Sports
Media and Arts	Animation
Media and Arts	Arts and Crafts
Media and Arts	Broadcast Media (TV, Radio)
Media and Arts	Media Production
Media and Arts	Motion Pictures and Film
Media and Arts	Music
Media and Arts	Newspapers and Magazines
Media and Arts	Online Media
Media and Arts	Photography
Media and Arts	Publishing
Media and Arts	Writing and Editing

glass.ai firm sectoral classification (2)

Sector Grouping	Sector
Non Profit and Education	Charities and Foundations
Non Profit and Education	E-Learning
Non Profit and Education	Education Management
Non Profit and Education	Higher Education and Universities
Non Profit and Education	Libraries
Non Profit and Education	Museums
Non Profit and Education	Places of Worship
Non Profit and Education	Primary and Secondary Education
Non Profit and Education	Professional and Trade Associations
Non Profit and Education	Professional Training and Coaching
Non Profit and Education	Think Tanks
Professional Services	Accounting
Professional Services	Business Supplies and Equipment
Professional Services	Consulting
Professional Services	Design
Professional Services	Information Services
Professional Services	Law Practice and Services
Professional Services	Market Research
Professional Services	Marketing and Advertising
Professional Services	Outsourcing and Offshoring
Professional Services	Public Relations and Communications
Professional Services	Security and Investigations
Professional Services	Staffing and Recruiting
Professional Services	Translation

Sector Grouping	Sector
Real Estate and Construction	Architecture and Planning
Real Estate and Construction	Building Materials
Real Estate and Construction	Civil Engineering
Real Estate and Construction	Construction
Real Estate and Construction	Executive Offices
Real Estate and Construction	Facilities Services
Real Estate and Construction	Glass, Ceramics and Concrete
Real Estate and Construction	Real Estate and Property Management
Supply Chain and Transport	Automotive
Supply Chain and Transport	Import and Export
Supply Chain and Transport	Logistics and Supply Chain
Supply Chain and Transport	Maritime
Supply Chain and Transport	Transportation, Trucking and Railroad
Supply Chain and Transport	Warehousing
Technology	Computer Games
Technology	Computer Hardware
Technology	Computer Networking and Security
Technology	Computer Software
Technology	Information Technology and Services
Technology	Internet
Technology	Semiconductors and Electronic Systems
Technology	Telecommunications and Wireless