SEEQUENT

GEOSCIENCE

DATA MANAGEMENT REPORT

**2021** and beyond

Contents

Foreword

Geoscience data underpins every major decision made on subsurface projects, from mining and energy exploration and production to infrastructure construction and environmental assessment.

Data Management is not simply about collecting and storing data. It sits at the core of business success – especially as resources become harder to find, and infrastructure projects become more complex.

It has been 10 years since the first Geoscience Data Management Report was published, and despite rapidly evolving practices and technology such as cloud computing, artificial intelligence and collaboration tools, consistent challenges remain. These challenges include ensuring data is securely and consistently stored and managed, while making it accessible, shareable, and auditable.

Companies who have begun to adopt cloud technology and new data management practices are slowly solving some of the old challenges, but new ones are emerging alongside them: how can data be used to optimise project costs, reduce risk and speed up production? How can we automate data analysis tasks to gain greater efficiency and certainty? How can machine learning and AI be applied to analyse data and become predictive?

While there are always going to be data challenges in geoscience, this report aims to examine and monitor what they are, how they are changing, and how different industries are responding, now and into the future.
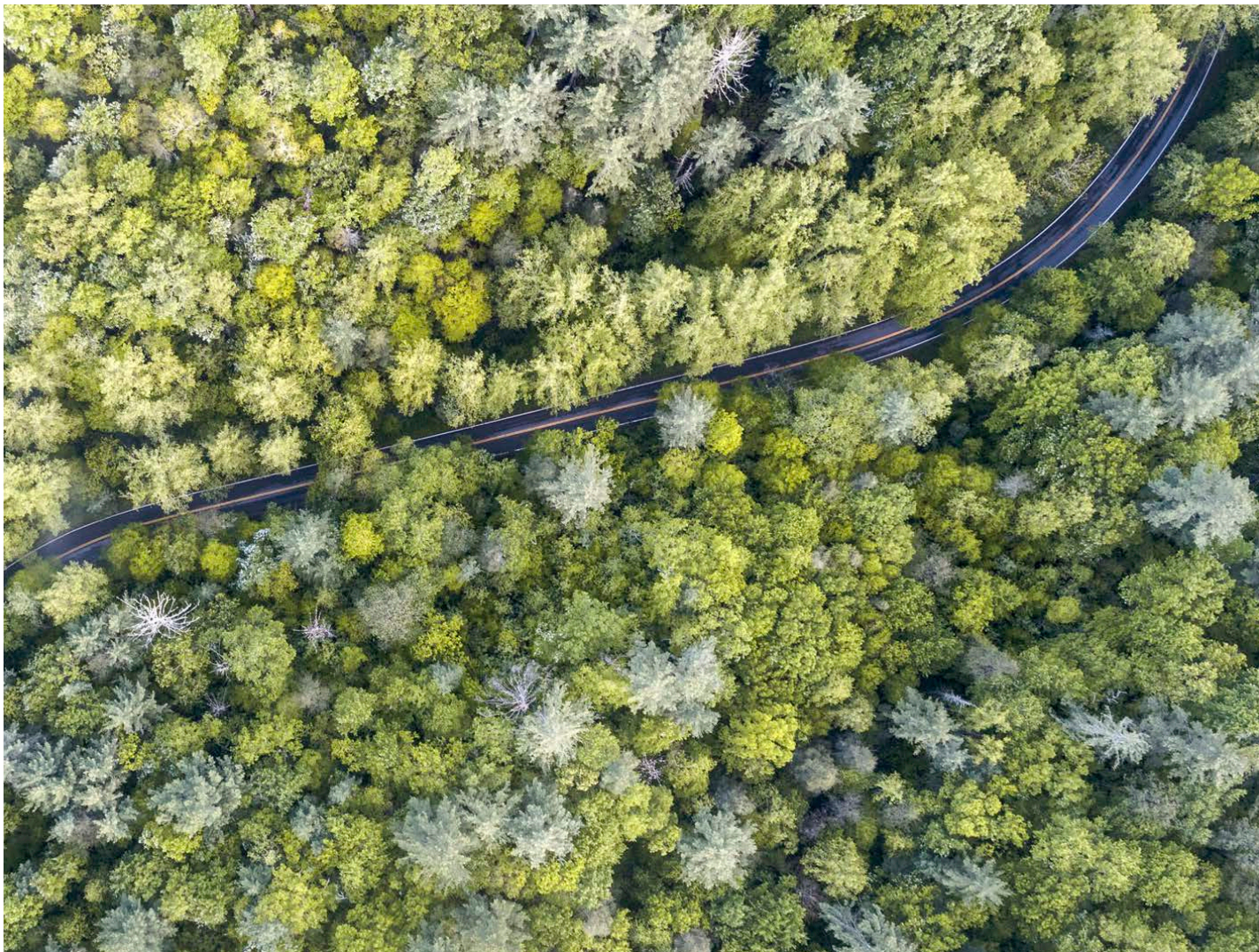
Executive Summary

The 2020 Geoscience Data Management Survey garnered 709 responses from 320 organisations. Geoscience data management continues to be a key issue across various industries. The survey identified key needs, challenges and successful approaches organisations are experiences. This Geoscience Data Management Report – 2021 and beyond, examines these findings.

83% of respondents stated that data management is a critical issue for their organisation.

Geoscience data management remains at least a top 5 issue for over 80% of organisations, yet less than a third currently use an established framework in their management processes. Since 2017 there has been a significant rise in the preference for commercial solutions.

Over 1 in 2 organisations are currently using or considering data practices such as data science scripting or machine learning/AI/advanced analytics.

The inclusion of these tools within software solutions will future proof them and provide opportunities to help businesses with growing volumes of data.

35% of respondents stated they spend greater than 30% of their time on data management tasks.

Over time there has not been a significant reduction in time spent on data management tasks and only 1 in 10 claim it is extremely easy to access data within their organisation. This suggests there is room for greater efficiency within data management by improving ease of access and reducing time spent on data management tasks.

40% of respondents stated version control was their biggest challenge when collaborating on active project data.

Confidence in managing data quality and quantity has remained consistent in the last 5 years which demonstrates there is still room for improvement. Confidence in data management is higher among those using a solution developed in house.

60% ranked data security and confidentiality of high importance

Data security and confidentiality remains the top data management challenge within organisations. This challenge ranks in the top 2 across all roles and industries but is less important for those in the environmental industry, where dependency on knowledge experts is most important.
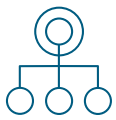
Since 2017, there has been a significant 12% increase in respondents wanting one point of access for data, and a single source of truth.

The top two desired outcomes for data management challenges are getting full value from data, and having one source of truth.

69% of respondents stated their readiness to leverage the cloud for geoscience data management. This is an increase of 14% since 2017.

Readiness to leverage the cloud for geoscience data management and access is up significantly from 2017. Although security is declining as a challenge it remains the top concern followed by performance (which is increasing as a challenge) and cost.

Introduction

This Geoscience Data Management Survey is the fifth iteration of its kind – the first report having been published in 2011. The 2020 survey ran across August and September 2020 and gaining a total of 709 responses across 320 organisations. A total of 16 questions were asked to gather insight across the following topics:

- Importance of geoscience data management
- Key challenges in geoscience data management
- Management of a breadth of geoscience data
- Barriers to geoscience data management success
- How geoscience organisations are overcoming their data management challenges
- Challenges faced when implementing a data management solution
- Level of readiness to leverage the Cloud for data management and access
- Main challenges when leveraging the Cloud for data management and access

Three new questions were added in 2020 which captured insight into geoscience organisations:

- Data management frameworks
- Ease of data access
- Data practices (i.e. data science scripting, advanced analytics, AI, Machine Learning)
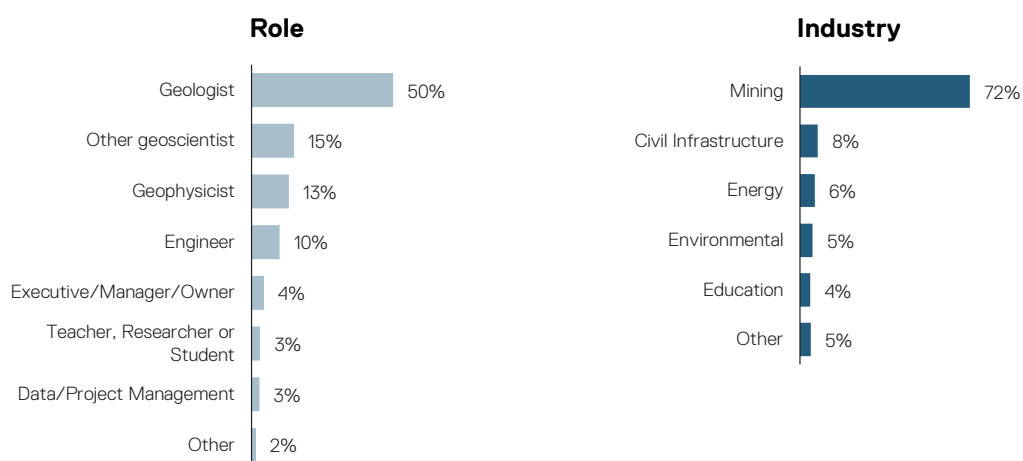
709
RESPONSES

91
COUNTRIES

320
ORGANISATIONS

Survey Statistics

Participants in the 2020 Geoscience Data Management Survey came from a broad range of countries and organisations.

The survey was distributed to a breadth of roles across mining, civil infrastructure, energy, environment and education. Geologists accounted for 50% of survey responses with an additional 28% working in geophysics or other geoscience roles. Engineers accounted for 10% of responses and the remaining 12% was represented by executive/managers/owners, teachers/researchers/students and data/project management or other roles.

Role

Geologist	50%
Other geoscientist	15%
Geophysicist	13%
Engineer	10%
Executive/Manager/Owner	4%
Teacher, Researcher or Student	3%
Data/Project Management	3%
Other	2%

Industry

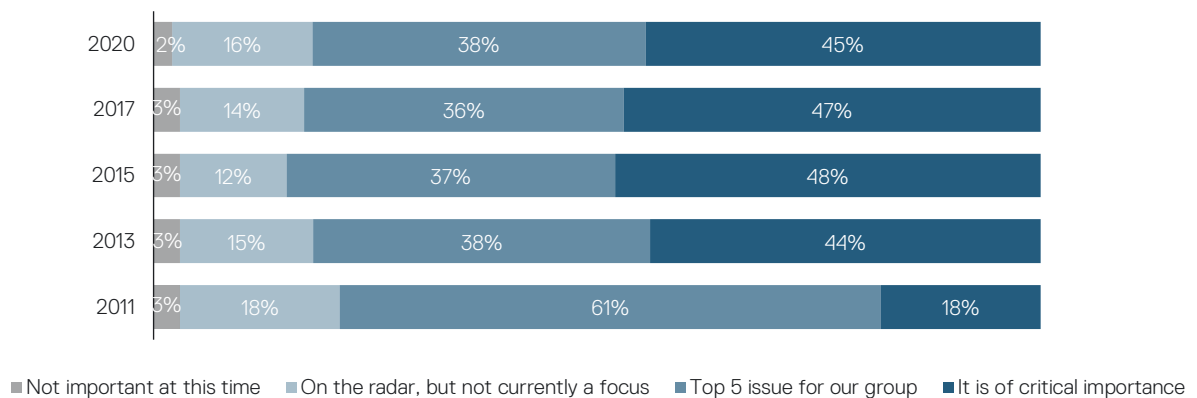|	
Mining	72%
Civil Infrastructure	8%
Energy	6%
Environmental	5%
Education	4%
Other	5%

Section 1

ROLE OF GEOSCIENCE MANAGEMENT

M anagement of geoscience data is vital for organisations as it allows for greater efficiency and extracting the most value from the data available. In 2020 the role of geoscience management was explored by understanding where data management ranks as an issue and identifying the importance of different data types within an organisation. Types of storage management, incidence of established frameworks and data practices were also covered.

Ranking of data management

Since tracking began, data management has been consistent with results seen since 2013 (see figure 1).
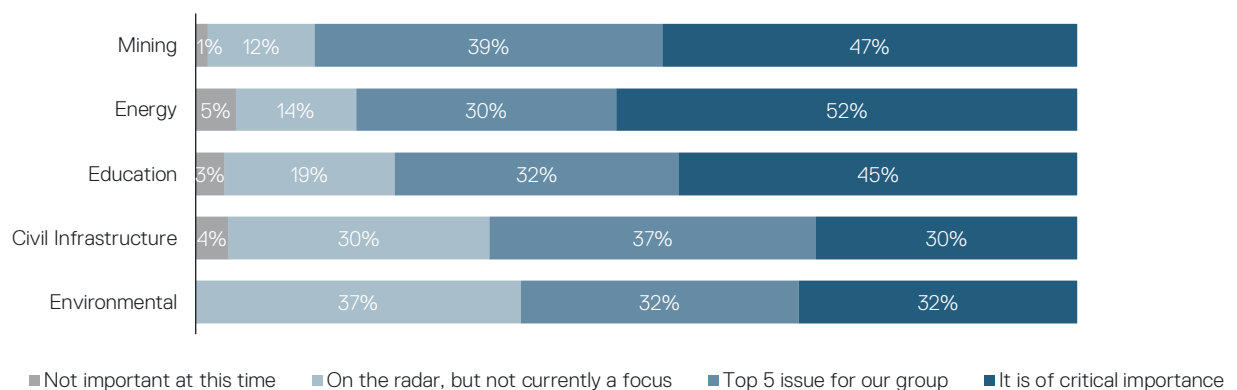
This issue is viewed differently among engineers vs. those in other job roles. Engineers are less inclined to view this as highly important, with 70% rating this issue as critical or top 5 compared to 83% among all other roles.

Figure 1. Ranking of Data Management as an Issue

Year	Not important at this time	On the radar, but not currently a focus	Top 5 issue for our group	It is of critical importance
2020	2%	16%	38%	45%
2017	3%	14%	36%	47%
2015	3%	12%	37%	48%
2013	3%	15%	38%	44%
2011	3%	18%	61%	18%

Differences also exist among industries with over 80% of those in mining and energy placing a high degree of importance on the issue. Conversely, those in civil infrastructure and environmental organisations place a lower degree of importance on the issue, with fewer than 70% rating it as critical or in the top 5 in these industries (see figure 2).

Figure 2. Ranking of Data Management by Industry

| Industry	Not important at this time	On the radar, but not currently a focus	Top 5 issue for our group	It is of critical importance
Mining	1%	12%	39%	47%
Energy	5%	14%	30%	52%
Education	3%	19%	32%	45%
Civil Infrastructure	4%	30%	37%	30%
Environmental		37%	32%	32%

Importance of data types

Across geoscience organisations there are numerous types of data being managed and several processes available for doing so. Respondents were asked to rate the importance of different types of data in their organisation and provide further context on their data management framework and storage/management system.

In 2020, respondents rank geological and drill hole or well as their most important (see Figure 3) and since 2017 there have been increases in importance seen for all but geophysical data (see Figure 4).

To be expected, the importance of data types varies by role and industry with 98% of geophysicists rating the importance of geophysical data as 4 or 5 out of 5 (top 2 box). Engineers place higher importance on geotechnical data (82% rate this as 4 or 5 out of 5) while geologists are more likely to rate geochemical data and drill hole or well data as important (top 2 box importance scores of 80% and 95% respectively).

Figure 3. Importance of Data Types
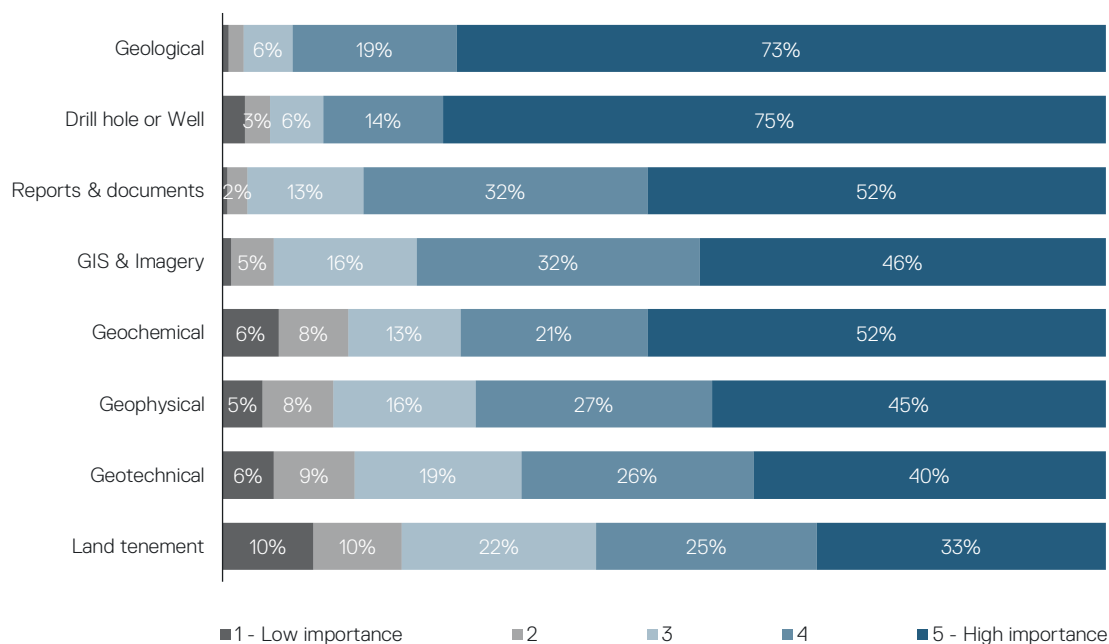
The importance of reports and documents has risen since 2017 (see Figure 4) and rates highly across the roles/industries.

Figure 4. Total Importance of Data Types Over Time (Top 2 box)
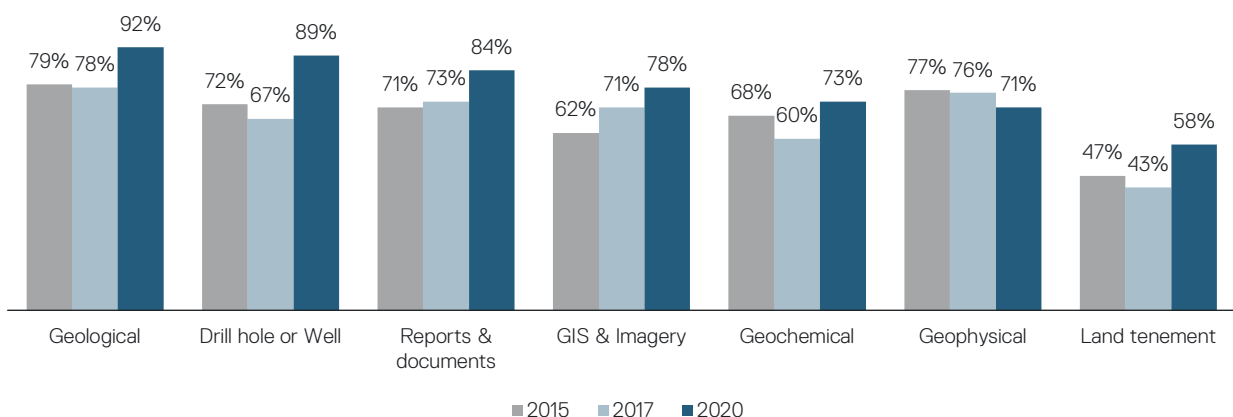
Complexity of data management frameworks

In a new question in 2020, respondents were asked to identify the type of data management framework used in their organisation. This revealed that that just under 1 in 4 have no defined data management framework (see figure 5).

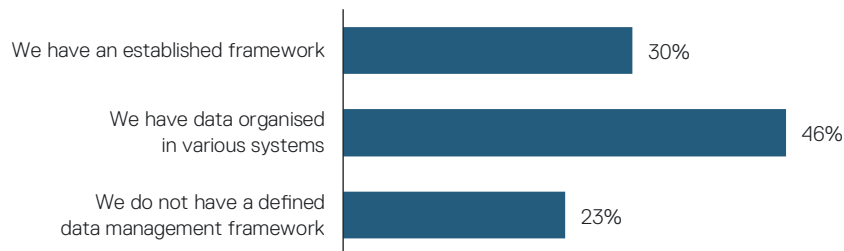Some differences existed within roles and industries, with the lack of a defined data management framework more common among engineers (38%) and those in education (52%). Established frameworks are more prevalent among geologists (34%) and those in the mining industry (34%).

Figure 5. Data Management Frameworks Used

Category	Percentage
We have an established framework	30%
We have data organised in various systems	46%
We do not have a defined data management framework	23%

How data is stored/managed within organisations

Consistent with findings from previous years, a centralised server is the most utilised approach for storage/management of geoscience data. The key difference in 2020 was the rise of commercial solutions which reached a peak of 18%, up significantly from 2015 and 2017 (see figure 6).

Reviewing these results by role and industry shows that geophysicists are less likely to be using commercial solutions (9%) while uptake is very small among those in education (3%) where the preference is for geoscientists to manage their own data (58%).

Figure 6. Storage/Management of Geoscience Data

	2013	2015	2017	2020
Folders/files on a centralised server	51%	52%	45%	43%
Geoscientists manage own data	20%	24%	29%	24%
Commercial solution	14%	8%	7%	18%
Internally developed solution	10%	13%	15%	13%
Other	6%	3%	4%	3%
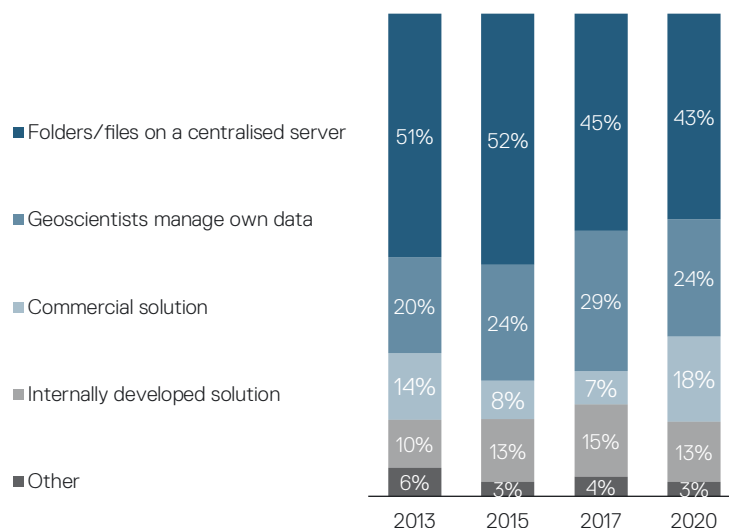
Data practices in use or under consideration

With an additional question in 2020, respondents were asked to indicate what data practices were currently used and/or under consideration in their organisation.

Over half are using or considering an enhanced data practice with just under a third using or considering data science scripting (see figure 7).

There is extensive crossover between these data practices with almost two thirds (62%) of those using data science scripting also using machine learning/AI and or artificial intelligence.

Usage or consideration of data science scripting is less prevalent among geologists (23%) and those in mining (26%).

Figure 7. Data Practices in Use/Under Consideration
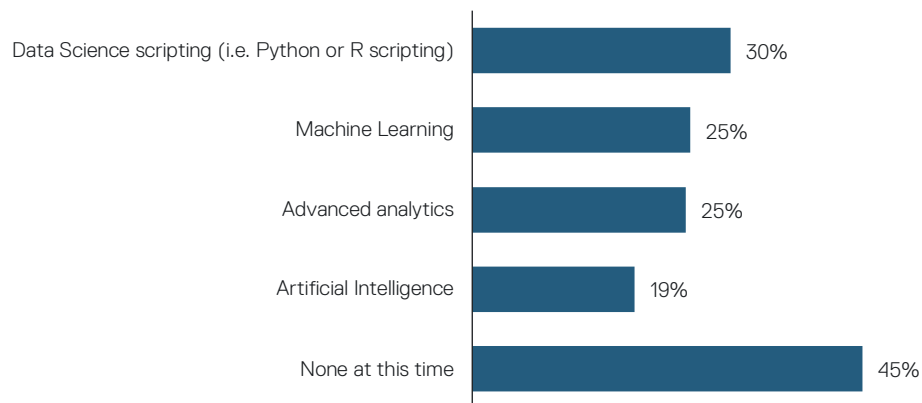
Practice	Percentage
Data Science scripting (i.e. Python or R scripting)	30%
Machine Learning	25%
Advanced analytics	25%
Artificial Intelligence	19%
None at this time	45%

Section 2

CHALLENGES FACED

The 2020 survey explored the challenges faced by geoscience organisations and how they handled the breadth of data within their organisation. Respondents were asked to rate their ease of data access, their level of confidence in handling data management challenges and challenges around collaboration. Time management was also explored to understand how much time was being dedicated to data management tasks.

Ease at which data can be accessed with organisations

A new question was implemented in 2020 which asked respondents to identify the ease at which they could access data within their organisation. Almost 1 in 2 (49%) feel that accessing data was easy with an equal amount experiencing at least some difficulty (see figure 8).

There is a link between the presence of an established data management framework and ease of data access; among those with an established framework, 67% find it easy to access data while this drops to 28% among those with no defined framework.

This finding is further supported when looking at results by roles and industry. Engineers and those working in education are more inclined to struggle with accessing data (top 2 box score 34% and 26% respectively), and these groups are also more likely to be operating without a defined framework.

Figure 8. Ease of Accessing Data Within Organisation

5%	10%	36%	38%	11%

■ 1 - Not at all easy   ■ 2   ■ 3   ■ 4   ■ 5 - Extremely easy
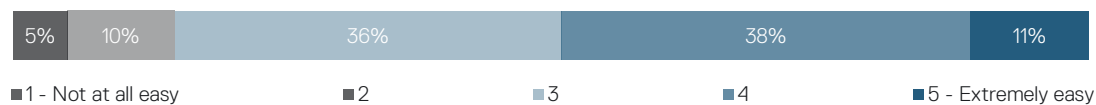
Time spent on data management tasks

In 2020, geoscientists are spending a considerable amount of time on data management tasks and this hasn't changed significantly from 2017. Time spent on data management tasks is not declining and little difference is seen irrespective of role, industry or data storage solution (see figure 9).

Figure 9. Percentage of Time Spent on Data Management Tasks

	2017	2020
More than 50%	15%	13%
Between 31 – 50%	21%	22%
Between 21 – 30%	23%	26%
Between 10 – 20%	19%	19%
Less than 10%	11%	10%
Unsure	11%	10%

■ 2017   ■ 2020
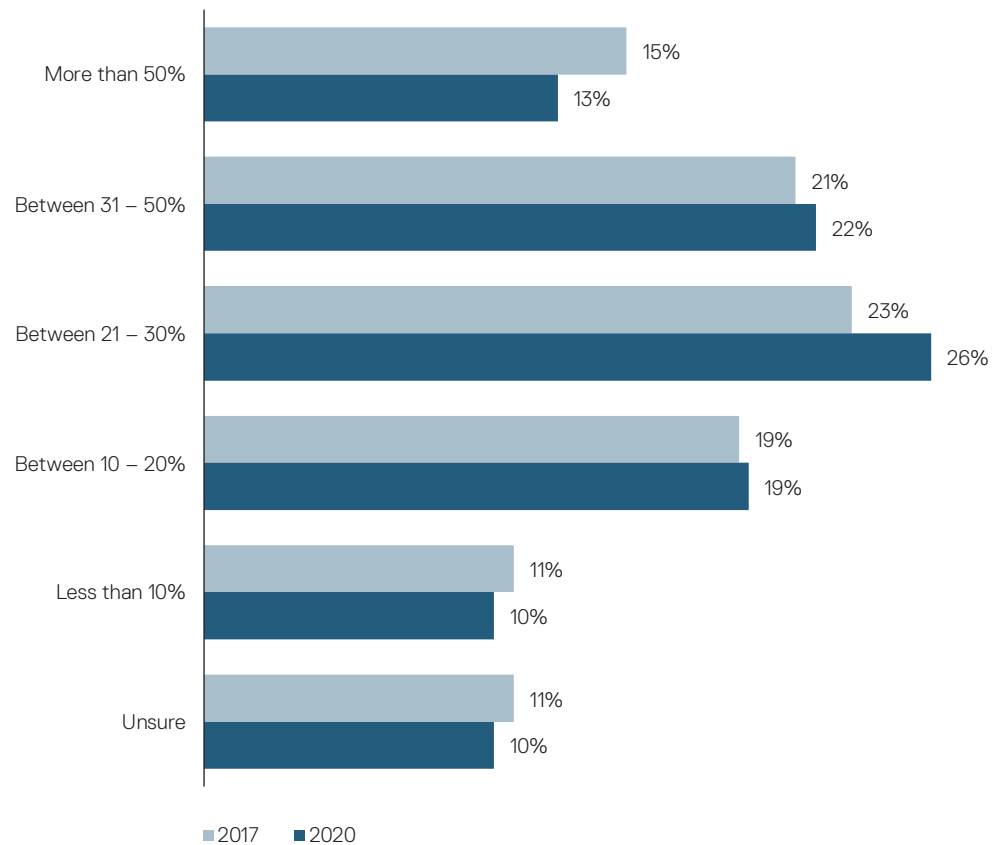
Level of confidence in handling different data management challenges

A high degree of importance is placed on data management, though there is still room for improvement. This is particularly relevant in raising confidence levels for managing data quantity and quality, neither of which have shown improvement since 2015 (see figure 11). Confidence in data duplication is improving but more work is required given less than half (44%) rate this highly (see figure 10).

Engineers have lower confidence in the quantity of data being managed, a finding that aligns with this group being more likely to have issues with ease of data access and lower utilisation of an established data framework. Within industries, it is those in civil infrastructure who have the lowest recorded confidence levels for quantity (43% top 2 box score), quality (46%) and data duplication (31%).

There is some evidence to suggest that in-house solutions are helping create greater confidence in relevancy, quality and quantity as these are significantly higher among those using this type of data management approach.

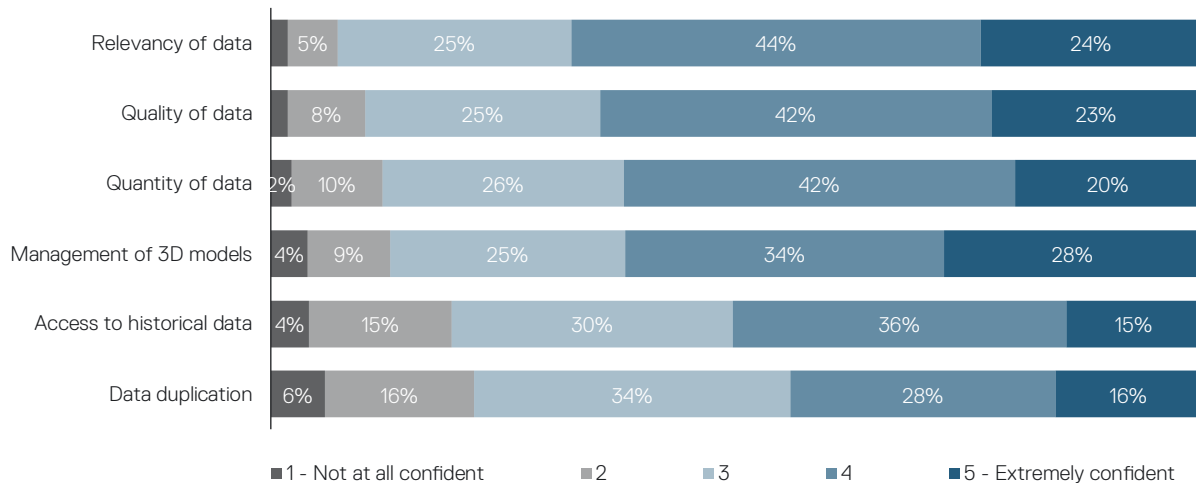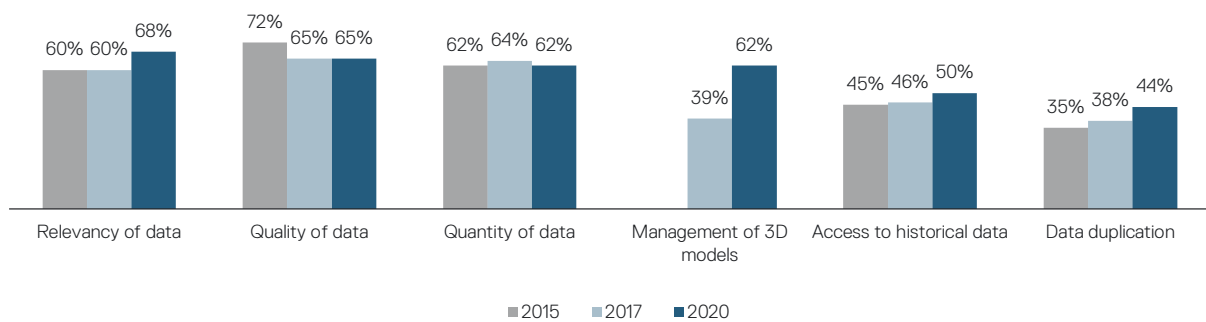**Figure 10. Confidence Levels in Handling Data Management Challenges**

	1 - Not at all confident	2	3	4	5 - Extremely confident
Relevancy of data		5%	25%	44%	24%
Quality of data		8%	25%	42%	23%
Quantity of data	2%	10%	26%	42%	20%
Management of 3D models	4%	9%	25%	34%	28%
Access to historical data	4%	15%	30%	36%	15%
Data duplication	6%	16%	34%	28%	16%

Figure 11. Total Confidence in Handling Data Management Challenges Over Time (Top 2 box)

	Relevancy of data	Quality of data	Quantity of data	Management of 3D models	Access to historical data	Data duplication
2015	60%	72%	62%	39%	45%	35%
2017	60%	65%	64%		46%	38%
2020	68%	65%	62%	62%	50%	44%

Challenges arising from collaborating on active project data

Collaboration presents two key challenges: ensuring most current/best quality version of the data and real time access in a useable format. These challenges have remained at the forefront since 2015 (see figure 12).

In 2020 these challenges do not differ significantly by role or industry, showing how inherent they are.

Figure 12. Challenges Collaborating on Active Project Data

- Knowing that you have the most current and best quality version of the data
- Real-time access to datasets in useable format
- Protecting datasets, restricting access (view/edit), ensuring security
- Reporting tools to see who has accessed/edited data
- Other

	2015	2017	2020
Knowing that you have the most current and best quality version of the data	35%	36%	40%
Real-time access to datasets in useable format	30%	35%	29%
Protecting datasets, restricting access (view/edit), ensuring security	12%	15%	18%
Reporting tools to see who has accessed/edited data	10%	10%	8%
Other	5%	4%	5%

Importance placed on data management challenges

Several key challenges were presented to respondents who then identified the level of importance for each.

In 2020, the most important data management challenge is data security and confidentiality which has grown in importance since 2017 (see figure 14). This challenge ranks in the top 2 across all roles and industries but is less important for those in the environmental industry, where dependency on knowledge experts is most important.

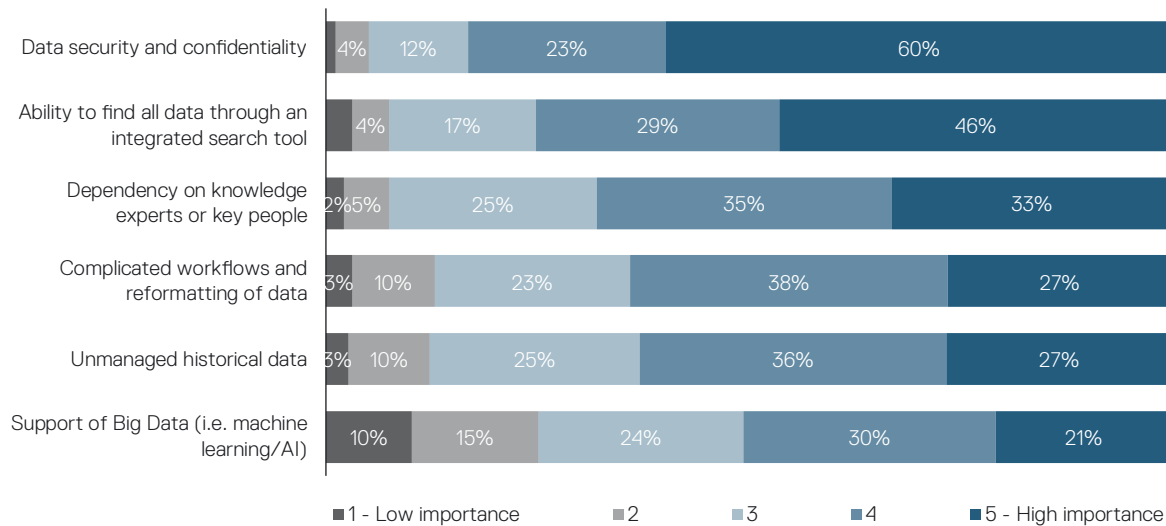**Figure 13. Importance of Data Management Challenges**

	1 - Low importance	2	3	4	5 - High importance
Data security and confidentiality		4%	12%	23%	60%
Ability to find all data through an integrated search tool		4%	17%	29%	46%
Dependency on knowledge experts or key people	2%	5%	25%	35%	33%
Complicated workflows and reformatting of data	3%	10%	23%	38%	27%
Unmanaged historical data	3%	10%	25%	36%	27%
Support of Big Data (i.e. machine learning/AI)	10%	15%	24%	30%	21%

Figure 14. Importance of Data Management Challenges Over Time
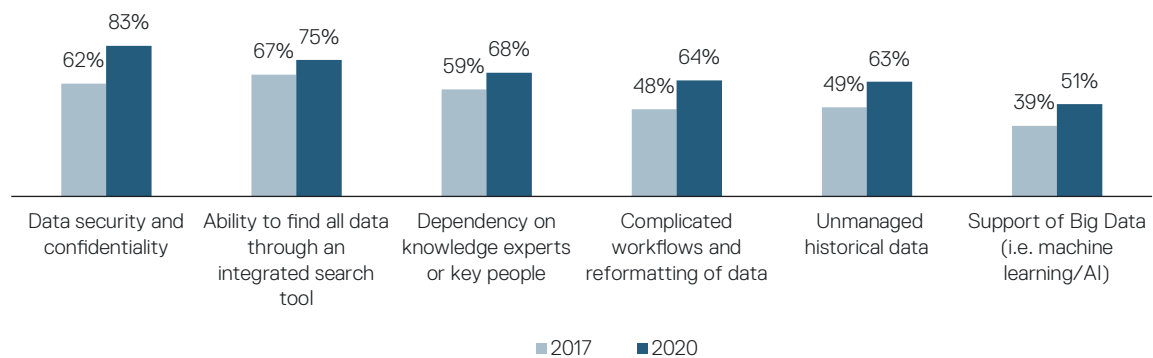
	2017	2020
Data security and confidentiality	62%	83%
Ability to find all data through an integrated search tool	67%	75%
Dependency on knowledge experts or key people	59%	68%
Complicated workflows and reformatting of data	48%	64%
Unmanaged historical data	49%	63%
Support of Big Data (i.e. machine learning/AI)	39%	51%

OVERCOMING DATA MANAGEMENT CHALLENGES

A s organisations face numerous challenges around data management, it is important to understand the key barriers to success from implementation through to maintenance and generating the desired outcomes.
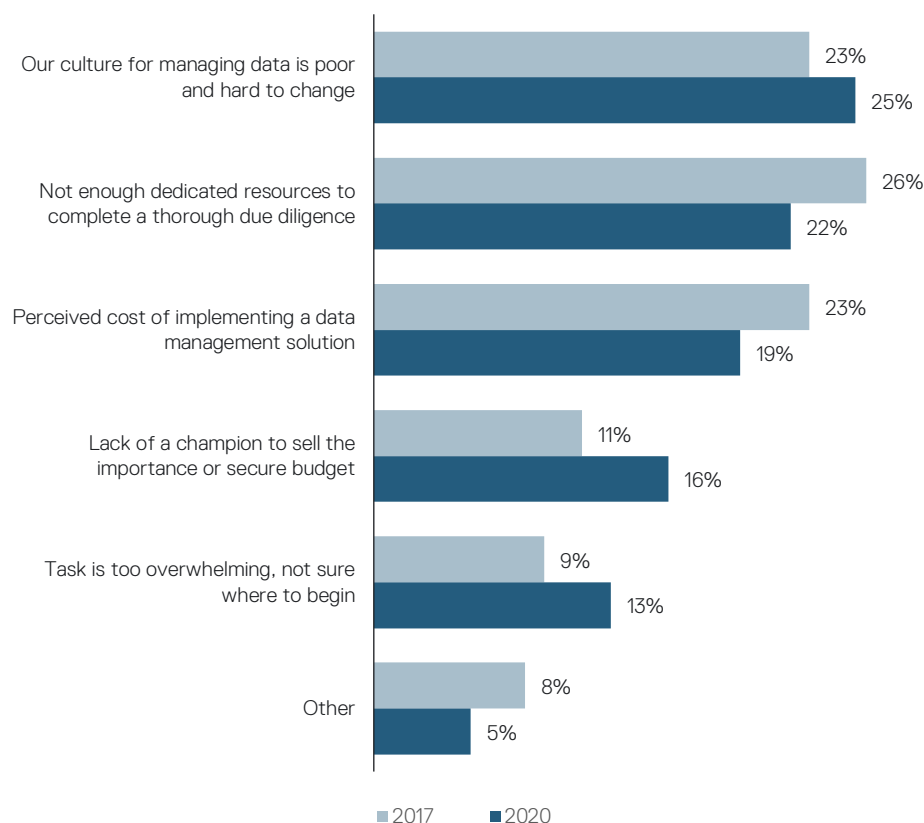
All these stages were explored in 2020 along with readiness to leverage the cloud and what challenges arise from using that technology.

Main challenge faced when implementing a data management solution

Following on from its addition in 2017, respondents were again asked to identify the main challenge faced when implementing a data management solution. In 2020 the challenges are still fragmented and centred on culture, resourcing and cost (see figure 15).

Few differences are evident among roles or industries, although for engineers the lack of a champion is more of a challenge (28%) and an established network is less prevalent within this role. Cost is more likely to be a challenge among those using a data management solution developed in house (30%).

Figure 15. Main Challenge in Implementing a Data Management Solution

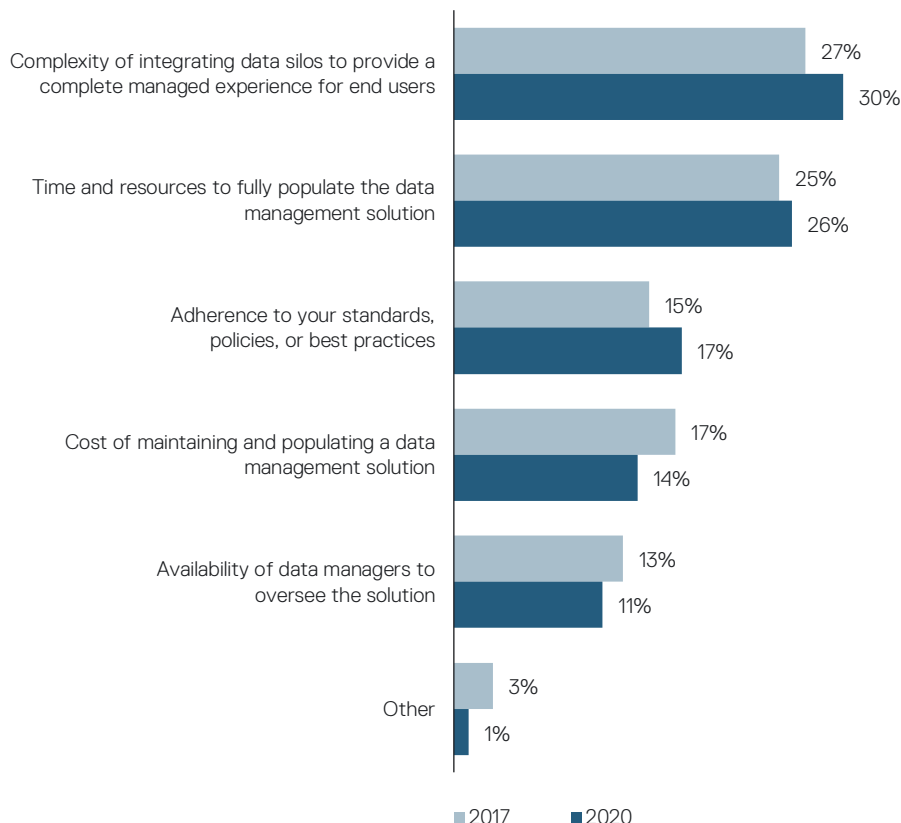|	2017	2020
Our culture for managing data is poor and hard to change	23%	25%
Not enough dedicated resources to complete a thorough due diligence	26%	22%
Perceived cost of implementing a data management solution	23%	19%
Lack of a champion to sell the importance or secure budget	11%	16%
Task is too overwhelming, not sure where to begin	9%	13%
Other	8%	5%

■2017 ■2020

Concerns regarding the maintenance and population of a data management solution

Complexity of data integration and time/resourcing remain the most important concerns held in relation to maintaining a data solution (30% and 26% respectively, see figure 16).

Complexity of data integration ranks high across most roles/industries but is less concerning to geophysicists (21%) and those in the Energy sector (11%), both of whom rate time/resources as their top concern. Cost is deemed to be less of a concern for those using a commercial solution (7%).

Figure 16. Most Important Concern When Maintaining a Data Solution

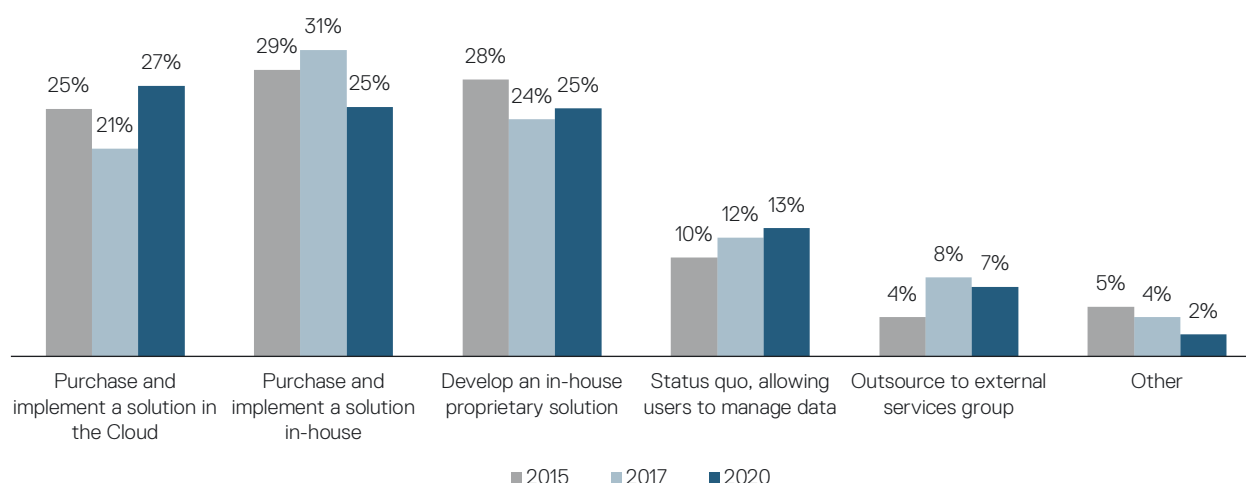|	2017	2020
Complexity of integrating data silos to provide a complete managed experience for end users	27%	30%
Time and resources to fully populate the data management solution	25%	26%
Adherence to your standards, policies, or best practices	15%	17%
Cost of maintaining and populating a data management solution	17%	14%
Availability of data managers to oversee the solution	13%	11%
Other	3%	1%

Preferred approaches to solving data management challenges

There are number of approaches available to organisations when solving data management challenges. Since 2017 there has been a shift in preference with cloud solutions now on the rise and in-house solutions in decline (see figure 17).

Cloud based solutions are a clear preference for those in civil infrastructure (39%), while developing a solution in-house ranks a clear number one for geophysicists (31%).

Figure 17. Approaches to Solving Data Management Challenges

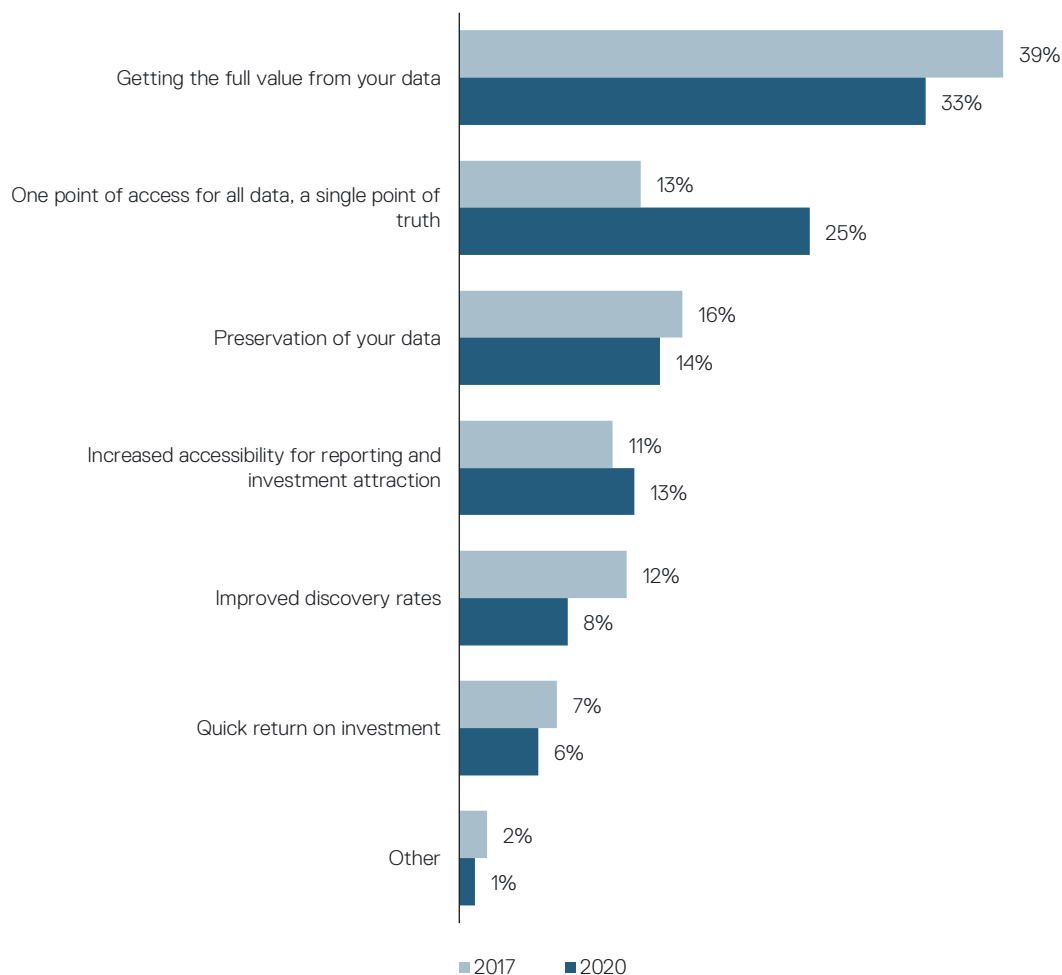|	2015	2017	2020
Purchase and implement a solution in the Cloud	25%	21%	27%
Purchase and implement a solution in-house	29%	31%	25%
Develop an in-house proprietary solution	28%	24%	25%
Status quo, allowing users to manage data	10%	12%	13%
Outsource to external services group	4%	8%	7%
Other	5%	4%	2%

Most important outcome expected from resolving data management issues

Getting the full value from data remains the most desirable, although less so in 2020. There has been a notable rise in those wanting a singular access point for their data, almost doubling from 13% in 2017 to 25% in 2020 (see figure 18).

Both outcomes are consistently placed in the top two across all roles and industries, signifying their widespread importance.

Figure 18. Desired Outcome From Resolving Data Management Issues

|	2017	2020
Getting the full value from your data	39%	33%
One point of access for all data, a single point of truth	13%	25%
Preservation of your data	16%	14%
Increased accessibility for reporting and investment attraction	11%	13%
Improved discovery rates	12%	8%
Quick return on investment	7%	6%
Other	2%	1%

Readiness to leverage the cloud and challenges this brings

Since 2017 respondents are significantly more likely to deem themselves ready to leverage the cloud, with the proportion feeling very/somewhat ready now at 69%, up from 55%. This change in attitude corresponds with the increased preference for cloud-based data management solutions (see figure 19).

Exploring the findings by role and industry, it is those in civil infrastructure who are most ready (74%), consistent with their preference towards cloud-based solutions for data management. However, level of readiness is strong across all industries and roles.

Figure 19. Readiness to Leverage the Cloud for Data Management and Access

	Unsure	Not ready	Somewhat ready	Very ready
2020	11%	20%	49%	20%
2017	12%	33%	38%	17%
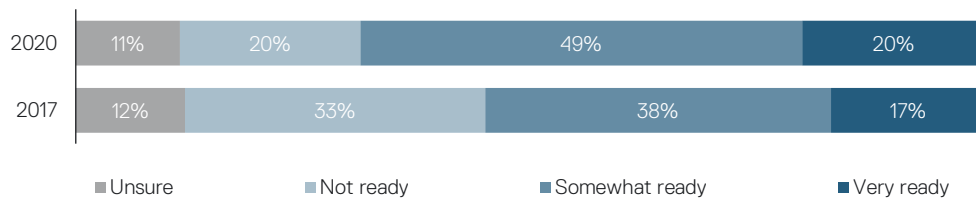
Security is the main concern in leveraging the cloud for data management, however since 2017 concerns for performance have risen while security has fallen (see figure 20).

Challenges differ by industry, with security ranking highest for mining and energy organisations, and cost ranking highest within education, civil infrastructure and environmental organisations (see Figure 21).

Numerous free text comments were captured, and these included a combination of options as well as concerns around internet connectivity which in part links back to potential issues around reliability.

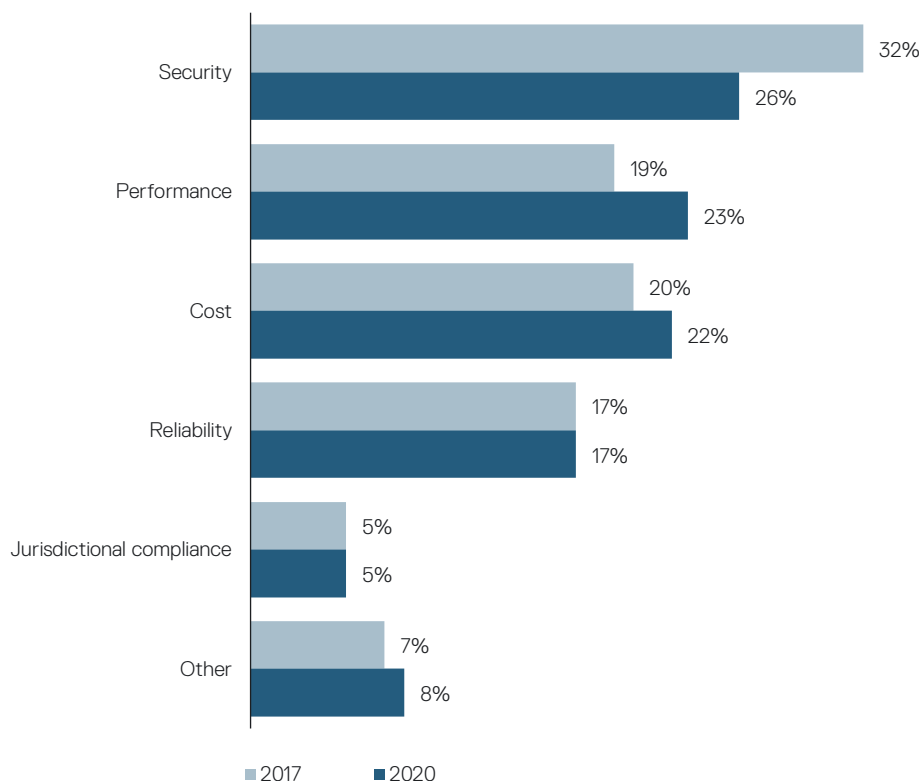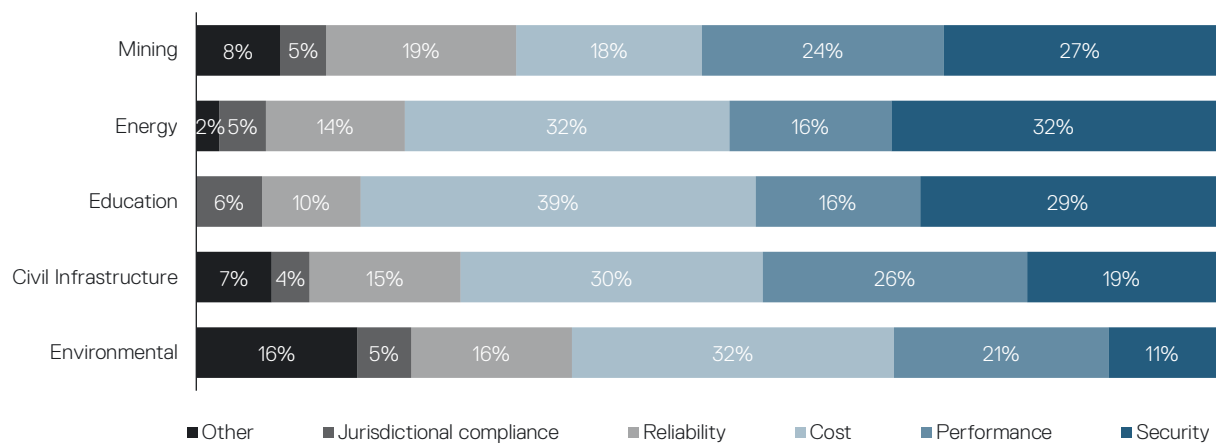**Figure 20. Challenges in Leveraging the Cloud for Data Management**

	2017	2020
Security	32%	26%
Performance	19%	23%
Cost	20%	22%
Reliability	17%	17%
Jurisdictional compliance	5%	5%
Other	7%	8%

Figure 21. Challenges in Leveraging the Cloud for Data Management by Industry

Industry	Other	Jurisdictional compliance	Reliability	Cost	Performance	Security
Mining	8%	5%	19%	18%	24%	27%
Energy	2%	5%	14%	32%	16%	32%
Education		6%	10%	39%	16%	29%
Civil Infrastructure	7%	4%	15%	30%	26%	19%
Environmental	16%	5%	16%	32%	21%	11%

■Other ■Jurisdictional compliance ■Reliability ■Cost ■Performance ■Security

Conclusion

Geoscience Data Management is a critical issue for organisations – a concern confirmed by 83% of our respondents across industries as varied as mining, civil infrastructure, energy, environment and education. Manage it well and organisations can drive their businesses, achieve greater efficiency and extract the maximum value from the data they gather (geological and drill hole data remaining the most important for our respondents).

However, almost 1 in 4 of the organisations we surveyed this year have no defined data management network to help them. Consequently geoscientists are still spending a considerable amount of time on data management tasks – a figure that is not improving (35% state they expend more than 30% of their time this way). Neither, it would seem, are confidence levels for managing data quantity and quality (no better than they were in 2015).

So what's preventing organisations from adopting better data management? The 2020 survey shows the challenges are fractured, encompassing culture, resources, cost and complexity. And while getting the full value from their data remains the primary objective for most organisations, the desire to have a single access point for it is growing rapidly – almost doubling since 2017.

The journey to better data management is an essential one, but in 2021 and beyond it remains a path that needs further focus to become faster, smoother and more effective – with benefits well recognised.
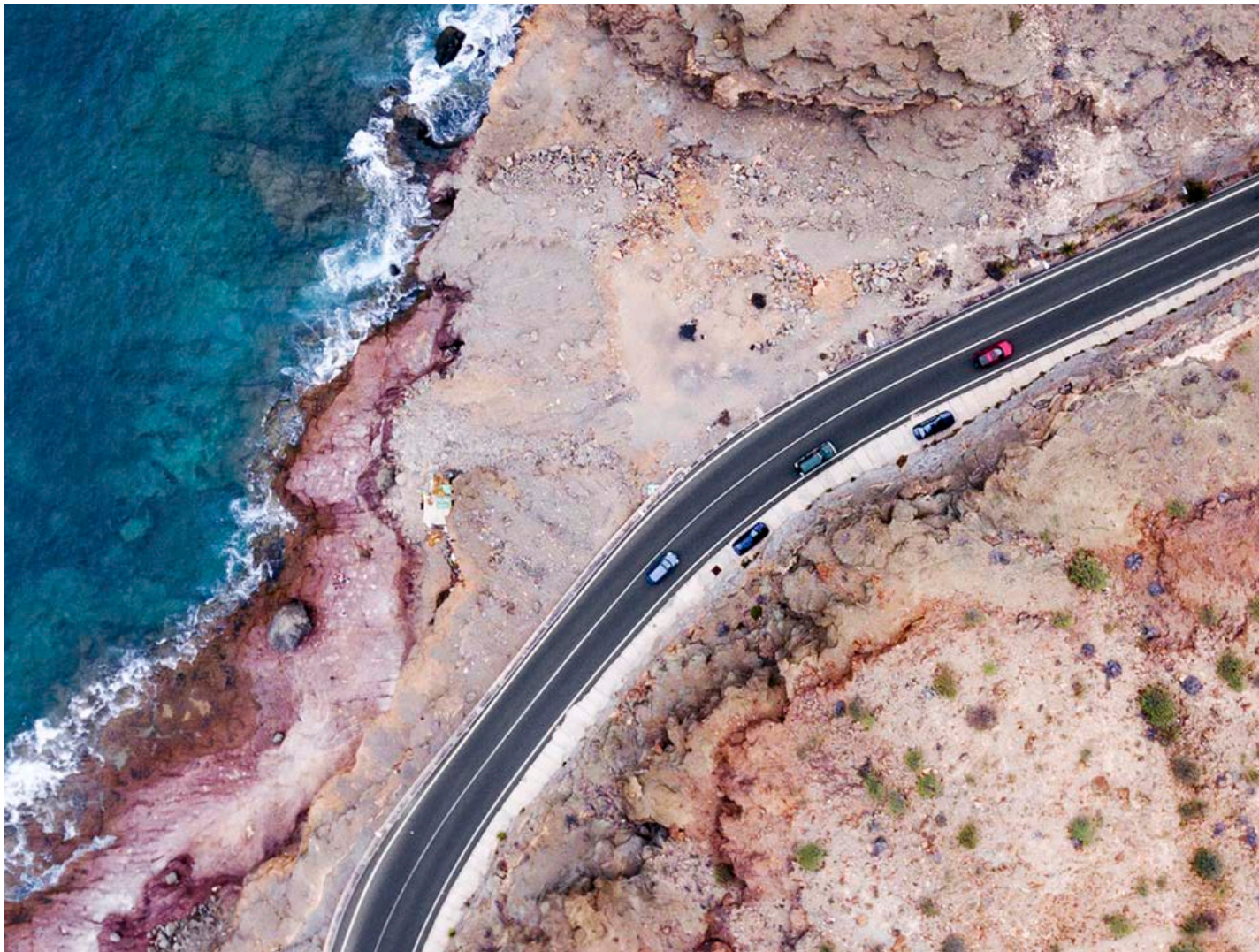
FROM COMPLEXITY TO CLARITY

Solve complex problems, manage risk and make better
decisions across the lifecycle of projects.

seequent.com