Gilder, C. E. L., Geach, M., Vardanega, P. J., Holcombe, E. A., & Nowak, P. (2021). Capturing the views of geoscientists on data sharing: A focus on the Geotechnical Community. *Quarterly Journal of Engineering Geology and Hydrogeology*, *54*(2), [qjegh2019-138]. https://doi.org/10.1144/qjegh2019-138

Date of version – 5th September 2020

Resubmitted to Quarterly Journal of Engineering Geology and Hydrogeology

Capturing the views of geoscientists on data sharing: A focus on the Geotechnical Community

Charlotte E. L. Gilder, MSci (Bristol) FGS
PhD student
University of Bristol
Department of Civil Engineering
Queen's Building
University Walk
Bristol, BS8 1TR, UK.
Email: charlotte.gilder@bristol.ac.uk
ORCID: 0000-0002-7695-5490

Martin Geach, PhD, BSc (Hons) CGeol FGS RoGEP (Professional)
Head of Information Management Services
Atkins,
Epsom, UK
Email: Martin.Geach@atkinsglobal.com

Paul J. Vardanega, BE MEngSc (QldUT) PhD (Cantab) GMICE MASCE MIEAust FHEA
Senior Lecturer in Civil Engineering
University of Bristol
Department of Civil Engineering
Queen's Building
University Walk
Bristol, BS8 1TR, UK.
Email: p.j.vardanega@bristol.ac.uk
ORCID: 0000-0001-7177-7851 (Corresponding Author)

Elizabeth A. Holcombe, MSci PhD (Bristol) GMICE FHEA
Senior Lecturer,
Department of Civil Engineering,
University of Bristol,
BS8 1TR,
United Kingdom.
Email: liz.holcombe@bristol.ac.uk
ORCID: 0000-0002-8517-6130

Paul Nowak, BSc (Lond.), ARSM, CEng, MICE, MIMMM, CGeol, FGS
Technical Director,
Atkins,
Epsom, UK
Email: paul.nowak@atkinsglobal.com

Capturing the views of geoscientists on data sharing: A focus on the Geotechnical Community

C. E. L. Gilder, M. Geach, P. J. Vardanega, E. A. Holcombe and P. Nowak

ABSTRACT

The sharing of Ground Investigation (GI) data within the United Kingdom (UK) is commonly practiced only in large infrastructure projects. A vast amount of GI data collected on routine projects is commonly not made publicly available which is arguably inefficient and potentially unsustainable. This paper captures the opinions of the geoscience community and the GI industry on data sharing to better understand current working practices and potential barriers to data sharing. The results of a survey carried out at the *Janet Watson Meeting 2018: A data Explosion: The Impact of Big Data in Geoscience* held at the Geological Society of London are reported. This survey is compared with the results of interviews undertaken during the Dig to Share project, a collaborative project led by Atkins, British Geological Survey (BGS) and Morgan Sindall. The opinions and practices of geoscientists towards data sharing across a project life cycle are reviewed. Drivers of risk relating to geotechnical aspects of a project are directly linked to current data sharing practice.

1. INTRODUCTION

The sharing of data and the use of digital tools are becoming ever more important aspects of delivery and management of large civil engineering projects in the United Kingdom (UK). Existing frameworks for collection and storage of ground investigation (GI) data (e.g., AGS 2004; BSI 2014; AGS 2017) can aid the process of providing geological and geotechnical data within the project team, allowing better decision-making, and making projects faster and more economical.

In a typical civil engineering project, a conceptual engineering geological model is created to anticipate what might be encountered onsite based on geological inferences (Fookes 1997; Parry et al. 2014; Norbury 2020). This is progressed to a preliminary observational model, one that is made from observations from available boreholes at a site (Parry et al. 2014). This preliminary work forms part of the Phase I - Desk Study (DS) (BSI 2020) and requires access to relevant historical data held in databases to help identify the potential presence of *technical* risks (e.g., Clayton 2001; McMahon 1985). A ground model includes engineering parameters, and a geotechnical model is built from a mathematical or physical analysis (Parry et al. 2014). Preliminary information, which would inform a preliminary observational model, may propagate through the subsequent design, construction, and service phases of a project in the form of geotechnical risks that will require identification and management. The DS is, therefore, widely recognised to be the most cost-effective part of this risk identification process (Figure 1) (Fookes 1997; Chapman, 2008; Griffiths 2014).

The quality of data collection is required to be consistent with the Geotechnical Category and specific to the requirements of the project. BSI (2020) requires a 'risk register' to be completed at the earliest point in a project to summarise the likely foreseen sources of unfavourable conditions associated with the subsurface. Geotechnical risks within a project can be described in terms of 'Technical' or 'Contractual' risks (Baynes 2010). The technical risks are those that are the result of geotechnical uncertainties, such as 'The risk of encountering an unknown geological condition'; 'The risk of using the wrong geotechnical criteria' and 'The risk of bias and/or variation in the design parameters being greater than estimated' (McMahon 1985). Geotechnical engineers usually manage these uncertainties and their associated risks using factors of safety and engineering judgement (often using limit state design or partial factoring approaches) (e.g., Simpson et al. 1981; Bolton 1981; Vardanega & Bolton 2016). Risk is also managed by development of documents such as the code of practice for ground investigations, BS 5930:2015+A1:2020 (BSI 2020) formally CP2001:1957, and BS EN 1997-2:2007 (Section 2) (BSI 2010) describing the careful planning required for collection of geotechnical data, or other specific documents such as Highways England (HE) document CD 622 (HE 2020) (superseding documents HD 22/08, BD 10/97 and HA 120/08).

This paper provides a brief overview of the progression of geotechnical data (and, therefore, the associated uncertainties and risks) through the early project phases in a typical UK construction context. The aim is to explore the current attitudes towards, and behaviour of, GI data storage and sharing of those working in the industry. The methodology and results of two studies (a semi-quantitative survey and a set of semi-structured qualitative interviews) are compared to understand: (i) to what extent does data sharing occur in practice? (ii) is open data useful? and (iii) does data sharing help to improve risk management? By tracking the path of GI data, along with the workflow of geotechnical engineers and geoscience professionals, the potentially significant role that data sharing could have in identifying geotechnical risks and potential for improvement of the risk reduction process is discussed.

2. REVIEW OF CURRENT UK GEOTECHNICAL DATA MANAGEMENT & WORKFLOW

Consideration of a centralised database for GI data came about as early as the 1980s. Organisations were urged to release the internal information they held at the time (Wood et al. 1982), and research became focused on the development of a suitable database structure relevant for descriptions of soils and rocks (e.g., Toll & Oliver 1995). The British Geological Survey (BGS) began collecting borehole records in the 19th century and this is maintained by the UK National Geoscience Data Centre (NGDC) (NGDC 2020). Data from ground investigations in the UK are often transferred within a project in a data transfer format file (AGS 2017) developed by the Association of Geotechnical Specialists (AGS) during the early 1990's and extended in various version to date (AGS version 4.1) (AGS 2020). The BGS provides, along with ‘pdf copies’ of groundwater wells and ground investigation boreholes, a web portal for upload and download of AGS files.

There is often a lack of distinction between geotechnical data and geotechnical information (Chandler et al. 2012). Geological *data* describes the factual data from a ground investigation whilst *information* describes the interpreted geological layers, 2D and 3D relationships (which includes pdf copies of borehole logs) (Chandler et al. 2012). In BS 8574:2014 (BSI 2014) the former distinction is termed ‘logical data’. Similarly, in the United States, a data transfer format known as the Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS) (DIGGS 2020), has been developed, but its adoption in practice has faced challenges. For instance, a study by the Ohio Department of Transport Office found that data is typically re-input multiple times by those performing ground investigation projects; the workflow for consultants typically involves manually re-typing data to produce pdf copies of information-based deliverables for reports (Deaton 2018).

Typical projects in the UK have a similar workflow as described in Deaton (2018) (Figure 2) and inefficiencies are caused by the way that data is handled during a project. The individuals closest to the original GI data at the start of the chain, i.e. new ground investigation data is produced at initiation of a project by the GI contractor, supply data down the chain to those that use the information to inform design decisions (see suggested groups of individuals in Figure 2). Where data is supplied as 'pdf copies', which is not logical data, the resulting work processes can cause significant delay in delivery of GI information. Geotechnical aspects can be submitted to Building Information Models (BIMs), yet these are commonly not reviewed, and the flow of information is often one way (Chandler et al. 2012). The data and information commonly do not reach a data repository, so cannot go towards benefitting another project team, researchers, or other organisations for improvement to UK geotechnical engineering practice.

Building Informational Modelling (BIM) (e.g., BSI 2019) government mandates have driven the need for collective development of 3-dimensional models into the wider project environment, which has affected geotechnical aspects (Chandler et al. 2012), and has driven the need to share data to build reliable ground models. An emphasis on 3D modelling also features in the BS 5930 amendment (BSI 2020; Norbury 2020). UK Government initiatives have led to the development of the BIM Industry Working Group to improve project delivery and operational performance (BIM Industry Working Group 2011; HMG 2013). In the Highways and Rail sectors in the UK, where there is a greater need to build detailed ground models (e.g., Mooney 2020), to achieve this requirement, organisations maintain databases of engineering related information acquired during infrastructure projects, i.e. the Highways Agency Geotechnical Data Management System (HA GDMS). This is a project-based system including scanned analogue (paper) reports and AGS files. Similarly, the United Kingdom Oil and Gas Authority (OGA), a government authority created to promote innovation in UK oil exploration, describes the need for stewardship of a National Data Repository of well and seismic information, in part due to new requirements of the UK Energy Act 2016. The geospatial commission was formed in April 2018 by the UK Government to co-ordinate driving of value from geospatial data (HMG 2019) in the context of construction and technology. This industry is looking to technologies including GIS, use of 'Big data' and BIM to promote innovation in geographical settings (AGI 2015).

In the construction industry greater efficiencies are seen in large infrastructure projects such as HS1 and HS2 (Smale 2017), Cross rail's Farringdon Station (Aldiss et al. 2012; Gakis et al. 2016) and modelling of the London Basin (Mathers et al. 2014) where data is shared within a project. These focused data sharing initiatives in UK infrastructure projects/sectors are presenting an interesting opportunity for the geotechnical community, as it is not currently understood to what degree an increase in data sharing would help the industry reduce project risk and impact overall productivity.

3. SURVEY METHODOLOGY

To investigate current attitudes to and practices of geoscience data-sharing a survey of 11 multiple-choice questions were presented to attendees of the *Janet Watson Meeting 2018: A data Explosion: The Impact of Big Data in Geoscience* held between 27th February to 1st March 2018 at the Geological Society of London, UK. A total 54 individuals responded to the questionnaire, of whom 44.4% represented the oil and gas industry, 26% research, 12.9% exploration, 7.4% construction industry, 3.7% remote sensing, 3.7% mining or mineral extractive industries and 1.8% data science (Q1). The purpose was to provide perspective of these individuals attitude towards ‘open data’ and so understand if the GI industry could improve their current workflow, by understanding the opinions of those working closer to data science concepts within the geoscience community. The questions were designed to first establish whether the individuals use open datasets, followed by an understanding of preferred data storage types, participation in release of data to open environments, how they use open data in their work and attitude to risk of using it.

To build on this initial research a comparison of the perspectives and practices of this wider geoscience community was made with those of a focussed population of construction industry professionals. A series of semi-structured interviews were held by the behavioural research initiative Dig to Share (2018a) a joint project between Atkins, Morgan Sindall, Fluxx and the BGS. This project produced a document which detailed interesting quotes from interviews of 23 people from the engineering sector (Dig to Share 2018b). Participants included individuals representing Utilities Providers, Ground Investigation/drilling Contractors, Multi-disciplinary Consultancies, Principal Contractors and Public-Sector Bodies. The interviews were designed with a particular focus on how these individuals interact with the existing BGS database. Participants were encouraged to discuss the following topics:

- value articulation: what is the perception of the value of the BGS database those dealing with ground investigation data on a day-to-day basis (additionally those who do not)?

- time and resources: is uploading data to the BGS database considered a commercially viable activity?
- complexity of parties involved in civil engineering projects: what are the problems around multiple stages of data sharing within a project?
- data ownership: should owners of the data be obliged to give over their data to open datasets?
- data availability and format: is the current data held in the BGS database at a high enough quality, quantity, accuracy to be useful? and
- technology: how can the functionality of the BGS database be improved, including the systems used by its contributors i.e. methods of data capture and manipulation tools?

This research collates the quotes from the Dig to Share document (Dig to Share 2018b) according to the participants stage on a typical project life cycle. The Janet Watson results are provided as a record of the number of answers for each multiple-choice question. The insight from both populations, the engineering sector, and from a wider geoscience context are reviewed.

4. RESEARCH INSIGHTS

4.1 Janet Watson Results

Figure 3 and Figure 4 provide the results of the Janet Watson meeting survey. The number of participants selecting each multiple-choice answer is provided. A total 48 out of 54 participants (88.8%) confirmed the use of open datasets in their work (Q2). The most common open resources selected from the choices shown included Ocean drilling/seismic based resources and both British and Non-UK Geological Survey data (Figure 3a). Internal/company held databases were the fourth most used resources. For the majority, open data is either extrapolated or used to support newly acquired data to make decisions (Figure 4a) and within sectors that are naturally more reliant on using shared data, such as Remote Sensing, Exploration and Oil and Gas, the use of open data as a sole informant for decisions appears to be an accepted approach. This acceptance of use of open data, is mirrored in the perception of risk, 73.1% of participants perceive a low, to moderately low attitude to risk (when risk is described as a scale proportioned across four categories i.e. low, moderately low, moderately high and high risk), 58% are from these sectors (Figure 4d). The small representation that are from roles which are either solely data science, or a mixture of data science and some other geoscience sectors, also held a lower perception of risk. 50% of participants felt comfortable asking a client for permission to make data collected during a project open source. The other 50% was made up of 25.9% who were not comfortable and 24.1% who were in a position where this was not applicable to them (Figure 3c).

Participants were asked their perceived level of sharing of geoscience data in their own work. The results indicated that participation to data sharing is low, in Figure 3d, 75% of survey participants believed less than 50% of data from their work is made available as open source. Two aspects which appear from results to be jointly contributing to low data sharing are the allocation of resources and time. The need to incentivise and prevent data loss is perceived to be of less importance. Also, it can be inferred that there is a general appreciation of financial advantages and benefits realised by the majority i.e. the lowest answered categories shown in Figure 3e. It is also evident that many geoscience professionals are still not prioritising data sharing activities in their workflow (Figure 4b). This question also identified that collaboration between data scientists and geoscientists is currently low. The participants were asked to provide an indication of who is most responsible for driving the release of data to open environments. A total 44% of the Janet Watson population agree that the drive is required from the Government (Figure 4c).

4.2 Dig to Share Analysis

The Dig to Share results compiled from document '*148 interesting things*' (Dig to Share 2018b) have been summarised into Figure 5 describing the key discussion themes drawn for comparison and arranged according to stage on a project life cycle. Allocations R-1 to R-6 have been used to show the groups of individuals (as identified in Figure 2) discussing each topic. The Dig to Share results indicate that generally, the BGS database is considered a valuable resource; a utilities provider using the words 'vital' or 'incredibly useful', whilst Multi-disciplinary consultants described advantage being gained to preliminary scoping and effective targeting of borehole logs, enabling preliminary work to be advised to designers. Interestingly those in roles closest to the data's original procurement (groups R-1 and R-2) are expressing most concern for the evaluation of uncertainties and, therefore, risk. From a GI Contractor at the beginning of the chain, there is a concern for the current quality of records held in the BGS database due to the lack of sharing of newly acquired data. This group also challenged the quality of meta-data due to improvement of testing methods or procedures, suggesting archived information needs to be checked, and describing difficulty in terms of resolution and quality when making 3D models. This group also indicated that open data will not be able to replace the need to complete a site-specific GI.

Those usually mid-way in the supply chain, the Multi-disciplinary consultants (R-2) provide discussion across all key themes. Concerns are regarding the processes surrounding quality control, i.e. the risk of accepting data that has an unknown audit process. Also describing a need to verify acquired data. This group noted the importance of ensuring reliability of data, and many discuss the liabilities associated with sharing data.

The differences in opinions across the supply chain represents an unbalanced view. Those at the infrastructure end consider the BGS open dataset a cost saving asset (Utilities provider), yet those earlier in the supply chain, see withholding data is a commercial advantage (Multi-disciplinary). One interviewee expresses a frustration of knowing data has been acquired during a nearby project, but they are unable to obtain it. Several participants of the Dig to Share interviews from Multi-disciplinary consultancy indicate that the act of sharing data is often forgotten, with no opportunity to bill for time. Various comments revolved around the cost burden that the additional time and work to share data would ensue, realising the impracticality of digitising data that remains in paper format or in archives (Infrastructure provider).

When discussing how communication through the supply chain is currently working, interviewees described the motivations and priorities of those that are the owners of the data. Several participants indicate that often the client or those working in the chain of GI to construction are unable to understand the value of the data that they may hold or own. To counteract this, others would prefer permission to become a common addition to contracts, enabling prior permission to be granted. Interestingly these comments are coming from the intermediary owners and users of the data, specifically the roles, R-2, R-3, and R-4.

When looking for future innovation to data sharing, those working on smaller scale developments indicate that planning authorities could help drive the open data agenda. Interviewees appear to be looking for solutions from existing work systems which have enabled other innovations in digital storage of data, i.e. BIM. Evidence of some public sector project requirements are described to already be written into contracts, for instance the Highways England contracts. Other interviewees indicate that those closest to the data, who are acquiring it, would be most suitable to deliver it to the BGS database. It is evident that existing use of technologies is currently low. Discussions include the preference to retain the system of capture of data in hand-written form onsite. Innovations in digital capture have been opposed by those who had experience testing these methods, due to concerns for possibility of missing the quality assurance associated with re-typing of logs by the geological or geotechnical engineer. Where an AGS digital data transfer format file is available the data is still chosen to be passed on as pdf copies.

5. DISCUSSION POINTS AND RECOMMENDATIONS

Figure 6 summarises the drivers of risk evidenced from the opinion studies. At project level disparity of meta-data resolution, low time and resources, perception of risk, participation and communication are identified. The wider geoscience community is affected by the lack of release of data beyond a project, which is inhibiting research opportunities and curation of learning outcomes sourced from industry experiences. This is further driving individual project risk, due to lack of innovation in research into codes of practices, natural variability and correlation of engineering parameters, which rely on data availability.

The drivers at project level are realised to be from 'human' sources, originating from current practices and understanding. It is clear from both survey populations that there is an appreciation of the benefits of data sharing but sharing participation is low. From those that find value in 'Big Data' principles and are working in roles that are dealing with finding solutions for our data needs in geosciences, the practical aspects of data sharing are the same, but the outlook of individuals are different. Each driver is considered separately in the following sections.

5.1 Time and Resources

The uploading of content and data sharing is not seen as commercially viable by the majority in the GI industry and it is not fully appreciated that the short-term efforts could outweigh that of the longer-term. Government strategies are already in place to drive reduction in construction costs and promote faster delivery (HMG 2013). Government data platforms are also detailing geographic information such as Lidar data, UK Air quality and the Department for Environment and Rural Affairs (DEFRA) which have been opened for use. A lot of the benefits are being realised in open data frameworks, developed to enable smart cities (mruk n.d.) and have government support (Capgemini Consulting 2013). It seems many want to benefit from the impacts of accessing more data but are still not allocating time and resources from within their own working systems.

The profession still encounters issues relating to the ground, which could have been identified prior to intrusive investigation and the scope or amount of geotechnical investigation is linked to the likelihood of under or overdesign (Jaksa et al. 2005). Whyte (1995) reviewed project expenditure on 58 UK highways projects reporting GI cost was 0.45 % of total out-turn cost. Where costs increased above tender cost, around 54% of the cost was due to geotechnical origins (Whyte 1995). Chapman (2008) reviews typical project expenditure from a developer's point of view and finds that the cost of a desk study (assumed £5K) and ground investigation (assumed £40K), typically constitutes 0.19% of the building cost, or 0.34% of the structure cost while ground risks are responsible for causing 20% of projects significant delay (where one month of delay is estimated to cost £100M). The continued pressure to deliver site investigations under low budgets may still be promoting competition between consultants and contractors.

Child et al. (2014) describe that the 'data journey' requires an update, which is more than just a digital version of the traditional paper process, where data can be accessed at all stages of a project development. Where a system for data sharing is initialised at the start of a project, this could help manage this aspect more efficiently.

5.2 Opportunities for dealing with risk

Risk was a common narrative discussed by those in the geotechnical sector and concerns are still held regarding liability. The multi-disciplinary consultants are often 'keepers' of the data and perhaps are most informed of its quality, understanding the difficulties that are held in communication of the data's aspects of uncertainty. Evidence from this study indicates that for other geoscience sectors who rely on shared data the use of open datasets as a sole source of information is accepted, and these individuals generally perceive a lower attitude to risk. However, the limit of liability is a complex issue in the construction sector, and one that needs to be better understood by all parties (e.g., AGS 2018). Problems of liability and risk are being addressed in other disciplines by using the concept of a 'data trust', as is described by the Open Data Institute (ODI) (ODI 2019). This may hold substantial benefits to those in construction. Additionally, new ways of writing contracts and project management approaches to attaining ground investigation data could easily be adopted (e.g., NGDC 2020, specifically 'Data sharing agreements'). To establish these new working methods the following points could be considered.

(1) Methods which allow an 'Agile' approach to management, which involve close collaboration with stakeholders of a project, so that unforeseen circumstances or anomalies found in the ground can be communicated and reviewed whilst the GI Contractor is still onsite so that risks are better understood (Geach & Grice 2020).

(2) The alignment of geotechnical outcomes to the principles of BIM with the progression of standard references to Levels of Detail and Levels of Information and/or procedures to enable quality control.

(3) Targeted efforts by supporting Learned Societies to increase understanding and dialogue between legal representatives and geotechnical professionals, such as accreditation and CPD training courses.

5.3 Collaboration and education

An outcome of this research is that that the owners of GI data and those in the supply chain must be better informed of technological advances in other sectors regarding data sharing and be aware of the existing frameworks that are currently not being effectively applied. In the GI sector, use of the *Code of Practice for the Management of Geotechnical data* (BSI 2014), and more recent modifications to AGS4 (AGS 2017; Child et al. 2014; AGS 2020), which include tables for the exchange of laboratory schedules within the format, could be more widely adopted. Griffiths (2014) describes a need for engineering geologists to broaden their skills base to tackle future societal challenges and it is clear that new technology needs to be further embraced by the community.

In the data science sector, many data management tools are already available which enable 'Big Data' application (Yaqoob et al. 2016). The use of 'Big Data' solutions in the engineering geoscience industry is reported to have increased, yet the limiting factor of effective uptake has similarly been described as human rather than technological (Dabson & Fitzgerald 2018). The current amounts of data in the GI industry are perhaps not as large as in other sectors, however, concepts such as PropBase (Kingdon et al. 2016), which can successfully combine data from differing formats could hold future benefits for integration of other geoscience datasets i.e. geophysical data, alongside the geotechnical properties held within an AGS data transfer format file. Other geoscience areas are also struggling to make the most of efficiency gains of technology. In the Petroleum and Petrochemical industry, the uptake of 'Big Data' is differing between upstream and downstream operations (Hassani & Silva 2018). In communities also dealing with geospatial data, it is understood that there is known financial and sustainability reward from sharing geographic information and this is actively encouraged but is also managing the same problems the GI industry is facing (e.g., AGI 2015; ODI 2018).

Consideration for data science collaboration is an important outcome of the Janet Watson meeting. The use of data science concepts and the enhanced methodologies for data analysis this can provide (i.e use of open source tools such as Python), should be given as much relevance as any other discipline taught to a geotechnical engineer. Interestingly, this is not discussed or considered an important competency in a lot of work seeking effective future development of engineering geology practice (i.e. Turner & Rengers 2010; Griffiths 2014).

The sharing and efficient storage of newly acquired data must form part of the workflow and be an integral part of the continued development of site investigation practice. Clear changes to existing practice are required, including:

a) Use of existing data science technology for management and querying of GI data (i.e. web-based systems, integration of analytical tools such as Python, R languages for visualisation and statistical analysis, using SQL or open-source data processing tools (i.e. Hadoop)) should be embraced by the industry. The AGS data transfer format should be the principle approach to sharing GI data within these developed systems.
b) Data sharing responsibilities should be nominated in contracts.
c) Increased employment from data science roles to introduce technological efficiencies in workflow systems, reducing pressures on time and resources.

5.4 Lack of communication and, therefore, feedback loop in the Geotechnical Community

The lack of movement to pro-active behaviours has been described previously about geotechnical engineering, describing some ground investigation practices as 'business open as usual' (Knill 2003). A large issue with the current workflow discussed in this research is communication. This is not only through lack of understanding by management or the design team (Bridges 2019), it is a lack of ownership or responsibility for data sharing described throughout the supply chain. The owners of the data are generally considered to be the ultimate client or developer of a particular asset. However, this is not clear in all projects and there is a need to specify data stewardship into contracts. The BGS is in the process of releasing donated site investigations and borehole logs which have been previously held as confidential, where the data has been held by them for over 4 years. This is in response to The Freedom of Information Act and Environmental Information Regulations (EIR) which requires BGS to revisit all previous confidentiality agreements and notifying donors (BGS 2020).

The Janet Watson results suggest that lack of sharing is not driven through reluctance to speak to owners of the data. The Dig to Share participants indicate that approaching the client is low on the list of priorities, teams can be large over a scheme and individuals cannot bill time to it. In response to these findings, the Dig to Share project developed the following initiatives which could improve upon communication, including:

a) Increase discussion in the value of geotechnical data with Clients who can benefit from data sharing principles.
b) Promote the role of Super Users (individuals within a company for instance) whose role is to drive an increase of data sharing from within a company.
c) Develop potential methods of incentivisation in the release and return of data for access to third parties in the national annex of borehole records (BGS database).

Janet Watson meeting participants identified both themselves and the government to be relevant drivers of change. Many participants from the GI industry believe that the changes are required from enforcement from government sources to drive the change in data sharing. There is already significant evidence of a government drive to the better management of geospatial data (HMG 2019). This is reflecting significant governmental investment, coupled with data mandates through principles of BIM.

Although there are clear benefits of sharing data it is evident that the geoscience community is still struggling to achieve this.

6. SUMMARY AND CONCLUSIONS

This paper draws together observations and views on the context of sharing data across the broader geoscience industry and more specifically the disciplines of ground investigation and geotechnical engineering. The questionnaire and interviews have provided an evidence base for current practices and some of the barriers to GI data sharing; and brought greater insight into an issue that has perhaps been recognised within the GI industry for many years. The main conclusions are: (i) data sharing is not an active part of the current workflow, and the current working practice for curation of data from past UK-based civil engineering projects is causing inefficiencies; (ii) in cases where data sharing is occurring it is providing useful and relevant information for preliminary phases of projects; and (iii) the lack of data-sharing is driving ground related uncertainties. Current working practices could be improved by actions such as increasing awareness of those in non-technical roles in the construction supply chain, allocating time and resources to data sharing, promoting data science in geoscience education, and increasing collaborations between data science professionals.

Data has historically been important for geotechnical engineering design processes and practices (e.g., Kulhawy & Mayne 1990; Phoon & Kulhway 1999) so it might be expected that the lack of data-sharing is hindering innovation. Research relies heavily on what has been published in the literature, by producing new data or through specific collaboration efforts which require funding (e.g., Vardanega et al. 2020). This is especially important given that ground-related uncertainties are still causing significant time and financial risks to projects. Concerning the management of geotechnical risks, it is proposed that data sharing not only hold potential for technical improvements and help inform project level management decisions, but additionally aid other projects and the continued research into geotechnical uncertainties in engineering design. The findings indicate that data sharing is not yet happening widely enough in the UK and two of the main barriers seem to be the current attitudes and working practices. The use of BGS as a central independent organisation to curate UK GI data is working. This suggests that continued collection and management needs to be fuelled both from company and government levels.

DATA AVALIABILITY STATEMENT

The numerical data related to the survey at the Janet Watson meeting are presented in the paper and anonymity is preserved. Data from the Dig to Share research can be sourced from Dig to Share (2018b).

ACKNOWLEDGMENTS

The first author would like to acknowledge the support the Engineering and Physical Sciences Research Council, Grant Number: EP/R51245X/1. The first author would also like to thank the Geological Society of London for supporting the authors undertaking of the survey at the Janet Watson Meeting. The authors thank the reviewers of the manuscript for their helpful comments which have helped improve the paper.

REFERENCES

Association for Geographic Information (AGI) 2015. *AGI Foresight Report 2020.* The Association for Geographic Information, London. UK. https://www.agi.org.uk/about/resources/category/100-foresight?download=160:agi-foresight-2020 [Accessed 05 September 2020]

AGS 2004. *Electronic Transfer of Geotechnical and Geoenvironmental Data, Edition 3.1.* Association of Geotechnical and Geoenvironmental Specialists, Kent, UK.

AGS 2017. *Electronic Transfer of Geotechnical and Geoenvironmental Data, Edition 4.0.4.* Association of Geotechnical and Geoenvironmental Specialists, Kent, UK.

AGS 2018. *LPA 68 – Guidance on Duty of Care arising from Third Party reliance on a geotechnical report: BDW Trading Ltd v Integral Geotechnique (Wales) Ltd [2018] EWHC 1915 (TCC).* Association of Geotechnical and Geoenvironmental Specialists, Kent, UK.

AGS 2020. *AGS Version 4.1*. Association of Geotechnical and Geoenvironmental Specialists, Kent, UK https://www.ags.org.uk/data-format/ags4-data-format/ [Accessed 05 September 2020]

Aldiss, D.T., Black, M.G., Entwisle, D.C., Page, D.P. and Terrington, R.L. 2012. Benefits of a 3D geological model for major tunnelling works: an example from Farringdon, east-central London, UK. *Quarterly Journal of Engineering Geology and Hydrogeology*, **45(4)**, 405–414. https://doi.org/10.1144/qjegh2011-066

Baynes, F.J. 2010. Sources of geotechnical risk. *Quarterly Journal of Engineering Geology and Hydrogeology*, **43(3)**, 321–331. https://doi.org/10.1144/1470-9236/08-003

BGS 2020. Site investigation and drilling information frequently asked questions. Depositing records and digital data. https://www.bgs.ac.uk/services/NGDC/records/depositing.html [Accessed 05 September 2020]

BIM Industry Working Group 2011. *A report for the Government Construction Client Group Building Information Modelling (BIM) Working Party Strategy Paper*. https://www.cdbb.cam.ac.uk/system/files/documents/BISBIMstrategyReport.pdf [Accessed 05 September 2020]

Bolton, M.D. 1981. Limit state design in geotechnical engineering. *Ground Engineering*, **14(6)**, 39–46.

Bridges, C. 2019. Geotechnical Risk: It's not only the ground. *Australian Geomechanics*, **54(1)**, 27–38.

British Standards Institution (BSI) 2010. *Eurocode 7: Geotechnical design – Part 2: Ground investigation and testing* (BS EN 1997-2: 2007+A1:2010). British Standards Institution, London, UK.

BSI 2014. *Code of practice for the management of geotechnical data for ground engineering projects* (BS 8574:2014). British Standards Institution, London, UK.

BSI 2019. *Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling – Part 2: Delivery phase of the assets* (BS EN ISO 19650-2:2018). British Standards Institution, London, UK.

BSI 2020. *Code of practice for ground investigations* (BS 5930:2015+A1:2020). British Standards Institution, London, UK.

Capgemini Consulting. 2013. *The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data*. https://www.capgemini.com/wp-content/uploads/2017/07/the_open_data_economy_unlocking_economic_value_by_opening_government_and_public_data.pdf [Accessed 05 September 2020]

Chandler, R.J., McGregor, I.D. and Morin, G.R. 2012. The role of geotechnical data in Building Information Modelling. In: *Proceedings of the 11th Australia New Zealand Conference on Geomechanics (ANZ 2012)*, 15-18 July 2012, Melbourne, Australia (Narsilio, G., Arulrajah, A. and Kodikara, J. (eds.)), 511–516. https://www.issmge.org/uploads/publications/89/82/2_1_10.pdf [Accessed 05 September 2020]

Chapman, T.J.P. 2008. The relevance of developer costs in geotechnical risk management. In: *Foundations: Proceedings of the Second BGA International Conference on Foundations*, ICOF2008 (Brown M.J., Bransby M.F., Brennan A.J. and Knappett J.A. (eds)). IHS BRE Press, 3-26.

Child, P., Grice, C. and Chandler, R. 2014. The Geotechnical Data Journey – How the Way We View Data is Being Transformed. *Information Technology in Geo-Engineering*, **3**, 83–88. https://doi.org/10.3233/978-1-61499-417-6-83

Clayton, C.R.I. 2001. Managing geotechnical risk: time for change? *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, **149(1)**, 3–11. https://doi.org/10.1680/geng.2001.149.1.3

Dabson, O. and Fitzgerald, R. 2018. Big Data in the industry: A critical examination of modern data collection and use in engineering geosciences. *In: Janet Watson Meeting 2018: Big Data in Geoscience Abstract Book* https://www.geolsoc.org.uk/~/media/shared/documents/events/Archive/2018/JW18%20Big%20Data%20abstract%20book.pdf [Accessed 05 September 2020]

Deaton, S.L. 2018. What are the Benefits of Geotechnical Data Interchange? Paper presented at: *69th Highway Geology Symposium (HGS)*. Proceedings available from: https://www.highwaygeologysymposium.org/wp-content/uploads/69_HGS-OPT.pdf [Accessed 05 September 2020]

DIGGS 2020. Data Interchange for Geotechnical and Geoenvironmental Specialists. http://www.diggsml.org [Accessed 05 September 2020]

Dig to Share 2018a. Digital Incubator, Infrastructure Industry Innovation Platform (i3P). Dig to Share. https://www.i3p.org.uk/projects/digtoshare/ [Accessed 05 September 2020]

Dig to Share 2018b. *148 interesting things we heard while interviewing 23 people in the engineering sector.* Atkins. Dig to Share. https://www.i3p.org.uk/wp-content/uploads/2018/09/03-148-Interesting-things...article.pdf [Accessed 05 September 2020]

Fookes, P.G. 1997. Geology for engineers: the Geological Model, Prediction and Performance. *Quarterly Journal of Engineering Geology*, **30(4)**, 293–424. https://doi.org/10.1144/GSL.QJEG.1997.030.P4.02

Gakis, A., Cabrero, P., Entwisle, D. and Kessler, H. 2016. 3D geological model of the completed Farringdon underground railway station. In: *Crossrail Project: Infrastructure, design and construction* (Black, M. (ed.)) Thomas Telford Limited and Crossrail, London, UK, **3**, 431-446.

Geach, M. and Grice, C. 2020. What is Agile Site Investigation and Why is Data so Vital? https://www.keynetix.cloud/technical-articles/agile-site-investigation/ [Accessed 05 September 2020]

Griffiths, J.S. 2014. Feet on the ground: engineering geology past, present and future. The 14th Glossop lecture. *Quarterly Journal of Engineering Geology and Hydrogeology*, **47(2)**, 116–143. https://doi.org/10.1144/qjegh2013-087

Hassani, H. and Silva, E.S. 2018. Big Data: a big opportunity for the petroleum and petrochemical industry. *OPEC Energy Review*, **42(1)**, 74–89. https://doi.org/10.1111/opec.12118

HE 2020. Managing geotechnical risk (CD 622 Revision 1). Highways England, UK. https://www.standardsforhighways.co.uk/prod/attachments/ff5ed991-71ed-4ff2-9800-094e18cd1c4c [Accessed 05 September 2020]

Her Majesty’s Government (HMG) 2013. Construction 2025. Industrial Strategy: government and industry in partnership. HM Government. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210099/bis-13-955-construction-2025-industrial-strategy.pdf [Accessed 05 September 2020]

HMG 2019. *Geospatial Commission Annual Plan 2019/2020.* HM Government. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/799197/6.5522-CO-GeospatialCommissionAnnualPlan.pdf [Accessed 05 September 2020]

Jaksa, M.B., Goldsworthy, J.S., Fenton, G.A., Kaggwa, W.S., Griffiths, D.V., Kuo, Y.L. and Poulos, H.G. 2005. Towards reliable and effective site investigations. *Géotechnique*, **55(2)**, 109–l21. https://doi.org/10.1680/geot.2005.55.2.109

Kingdon, A., Nayembil, M.L., Richardson, A.E. and Smith, A.G. 2016. A geodata warehouse: Using denormalisation techniques as a tool for delivering spatially enabled integrated geological information to geologists. *Computers and Geosciences*, **96**, 87–97. https://doi.org/10.1016/j.cageo.2016.07.016

Knill, J. 2003. Core values: the first Hans-Cloos lecture. *Bulletin of Engineering Geology and the Environment*, **62(1)**, 1–34. https://doi.org/10.1007/s10064-002-0187-9

Kulhawy, F.H. and Mayne, P.W. 1990. Manual on Estimating Soil Properties for Foundation Design. *Rep. No. EL-6800*, Electric Power Research Institute, Palo Alto, CA, USA.

Mathers, S.J., Burke, H.F., Terrington, R.L., Thorpe, S., Dearden, R.A., Williamson, J.P. and Ford, J.R. 2014. A geological model of London and the Thames Valley, southeast England. *Proceedings of the Geologists’ Association*, **125(4)**, 373–382. https://doi.org/10.1016/j.pgeola.2014.09.001

McMahon, B.K. 1985. Geotechnical Design in the face of uncertainty. *Australian Geomechanics Journal*, **10(1)**, 7–19.

Mooney, N. 2020. Ameys’ Story: Smart use of Historical Geotechnical Data for a Smart Motorway. https://www.keynetix.cloud/technical-articles/historical-geotechnical-data/ [Accessed 05 September 2020]

mruk n.d. Building A Future City: Future City Glasgow Evaluation. Glasgow City Council. https://futurecity.glasgow.gov.uk/reports/12826M_FutureCityGlasgow_Evaluation_Final_v10.0.pdf [Accessed 05 September 2020]

NGDC 2020. *Deposit data with NGDC.* National Geoscience Data Centre, British Geological Survey. https://www.bgs.ac.uk/services/ngdc/guidelines.html [Accessed 05 September 2020]

Norbury, D. 2020. Ground models – a brief overview. *Quarterly Journal of Engineering Geology and Hydrogeology*, https://doi.org/10.1144/qjegh2020-018

ODI 2018. The UK’s geospatial data infrastructure: challenges and opportunities. *Open Data Institute*. https://theodi.org/wp-content/uploads/2018/11/2018-11-ODI-Geospatial-data-infrastructure-paper.pdf [Accessed 05 September 2020]

ODI 2019. Data Trust for the Royal Borough of Greenwich and Greater London Authority. Pilot 1 – Sharing Cities Transport and Energy Case Studies. Open Data Institute. http://theodi.org/wp-content/uploads/2019/04/BPE_PITCH_GREENWICH_GLA_A4-FINAL.pdf [Accessed 05 September 2020]

Parry, S., Baynes, F.J., Culshaw, M.G., Eggers, M., Keaton, J.F., Lentfer, K., Novotny, J. and Paul, D. 2014. Engineering geological models: An introduction: IAEG commission 25. *Bulletin of Engineering Geology and the Environment*, **73(3)**, 689–706. https://doi.org/10.1007/s10064-014-0576-x

Phoon, K-K. and Kulhawy, F.H. 1999. Characterization of geotechnical variability. *Canadian Geotechnical Journal*, **36(4)**, 612–624. https://doi.org/10.1139/t99-038

Simpson, B., Pappin, J.W. and Croft, D.D. 1981. An approach to limit state calculations in geotechnics. *Ground Engineering*, **14(6)**, 21-26, 28.

Smale, K. 2017. Sharing geotechnical information 'could cut costs'. *New Civil Engineer*. https://www.newcivilengineer.com/business-culture/sharing-geotechnical-information-could-cut-costs/10025179.article [Accessed 05 September 2020]

Toll, D.G. and Oliver, A.J. 1995. Structuring soil and rock descriptions for storage in geotechnical databases. *Geological Data Management, Geological Society Special Publication*, **97(1)**, 65–71. https://doi.org/10.1144/GSL.SP.1995.097.01.08

Turner, A.K. and Rengers, N. 2010. A Report Proposing the Adaptation of the ASCE Body of Knowledge Competency-based Approach to the Assessment of Education and Training Needs in Geo-Engineering. Progress report to the: Joint Technical Committee JTC-3: Education and Training. https://www.iaeg.info/wp-content/uploads/2019/01/c4_jtc3-appendix-1-turnersandrengers-jan2010.pdf [Accessed 05 September 2020]

Vardanega, P.J. and Bolton, M.D. 2016. Design of Geostructural Systems. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, **2(1)**, [04015017] https://doi.org/10.1061/AJRUA6.0000849

Vardanega, P.J., Crispin, J.J., Gilder, C.E.L., Voyagaki, E. and Ntassiou, K. 2020. DINGO: A Pile Load Test Database. *Piling 2020* (to be presented).

Whyte, I.L. 1995. The financial benefit from a site investigation strategy. *Ground Engineering*, **28**, October, 33–36.

Wood, L.A., Tucker, E.V. and Day, R. 1982. Geoshare: The development of a databank of geological records. *Advances in Engineering Software*, **4(4)**, 136–142. https://doi.org/10.1016/0141-1195(82)90015-8

Yaqoob, I., Hashem, I.A.T, Gani, A., Mokhtar, S., Ahmed, E., Anuar, N.B. and Vasilakos, V. 2018. Big data: From beginning to future. *International Journal of Information Management*, **36(6B)**, 1231–1247. https://doi.org/10.1016/j.ijinfomgt.2016.07.009

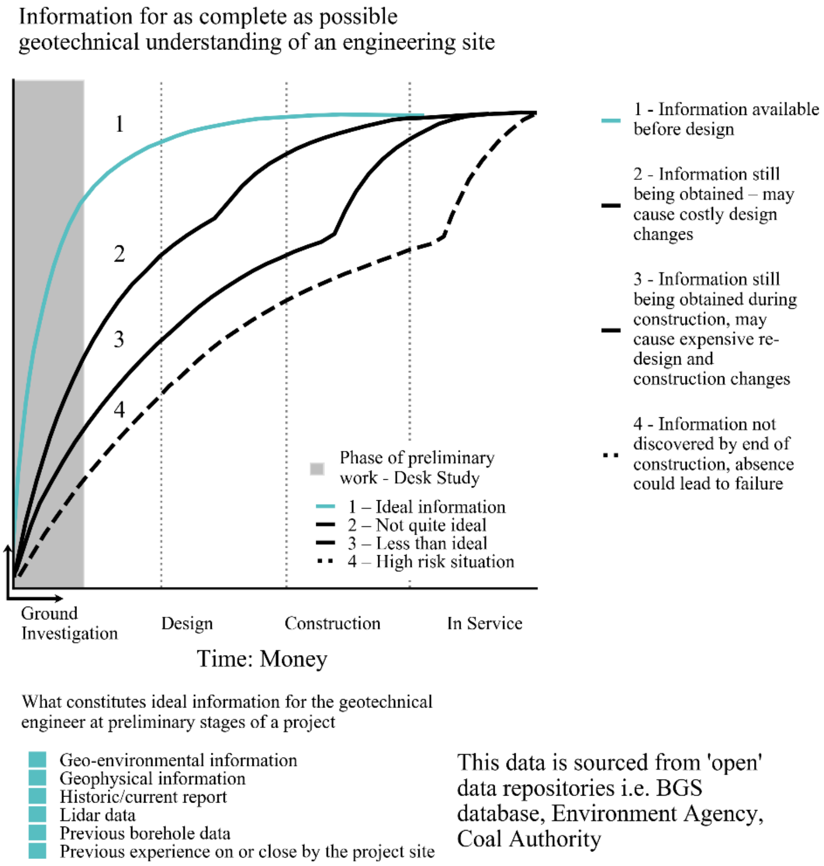

Figure 1. Modified from Fookes (1997) schematic. Relationship between time and money spent on a civil engineering project and the influence of information gathering to enable a complete geotechnical understanding of a project. © Geological Society of London

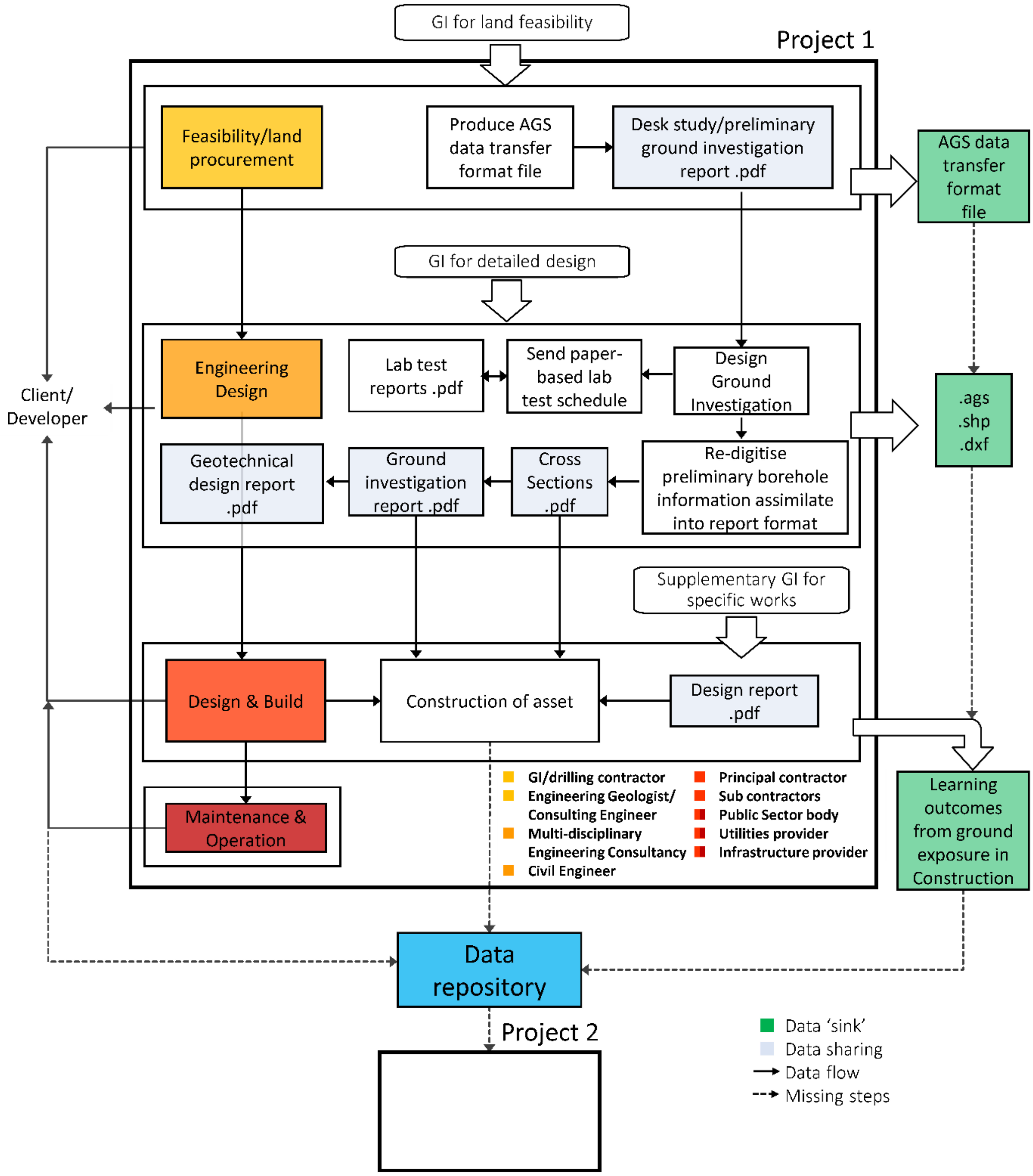

Figure 2. Summary of the main stages of ground investigation work undertaken during a civil engineering project including the data transferred and the professional roles involved in each stage of work.

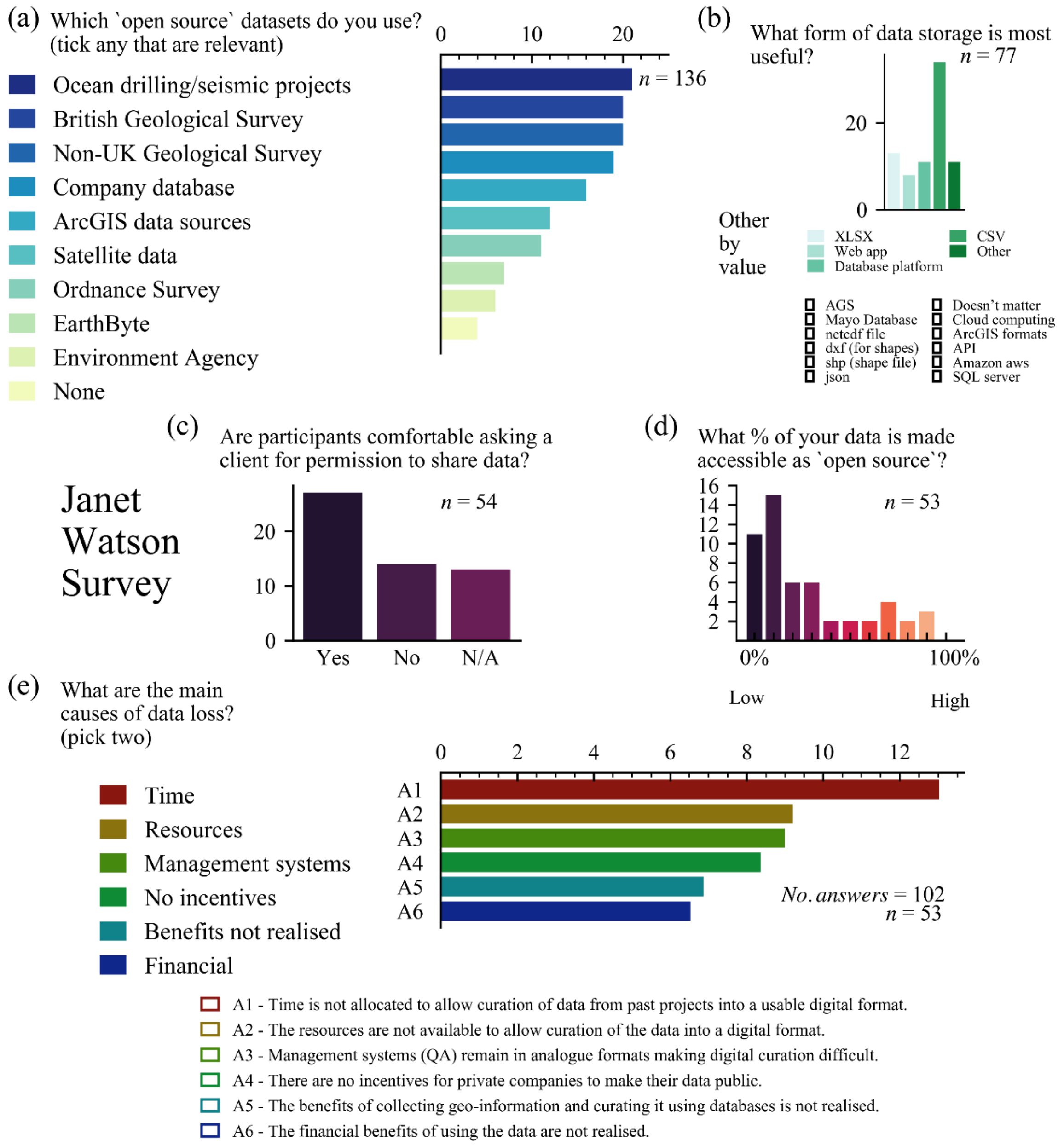

Figure 3. Results from Janet Watson Survey providing the number of answers for each question. Questions are as follows: (a) which open datasets do you use? (Q3) Ocean drilling/seismic projects includes seismic data and sources relevant to oil exploration including: International Ocean Discovery Program (IODP), Deep Sea Drilling Project (DSDP). Non-UK Geological Surveys include United States Geological Survey (USGS) and Geoscience Australia. Other participants suggested the following additions to the original multiple-choice options presented above including: Geological Survey of the Netherlands (TNO), Common Data Access Ltd (CDA) and Norwegian Petroleum Directorate (NPD) (b) What form of data storage is most useful? (Q5) additional suggestions by participants are added (c) Are participants comfortable asking a Client for permission to make the data collected on their project to be made open source? (Q9) (d) What percentage of your data is made accessible as open source? (Q7) 0% = none of their work is perceived to be shared 100% = all of their work is perceived to be shared. (e) What is the main cause of data loss in your industry? No of answers shows where some participants did not pick two, each participant answer was weighted so that the total score per participant was equal to one (one participant did not answer) (Q6).

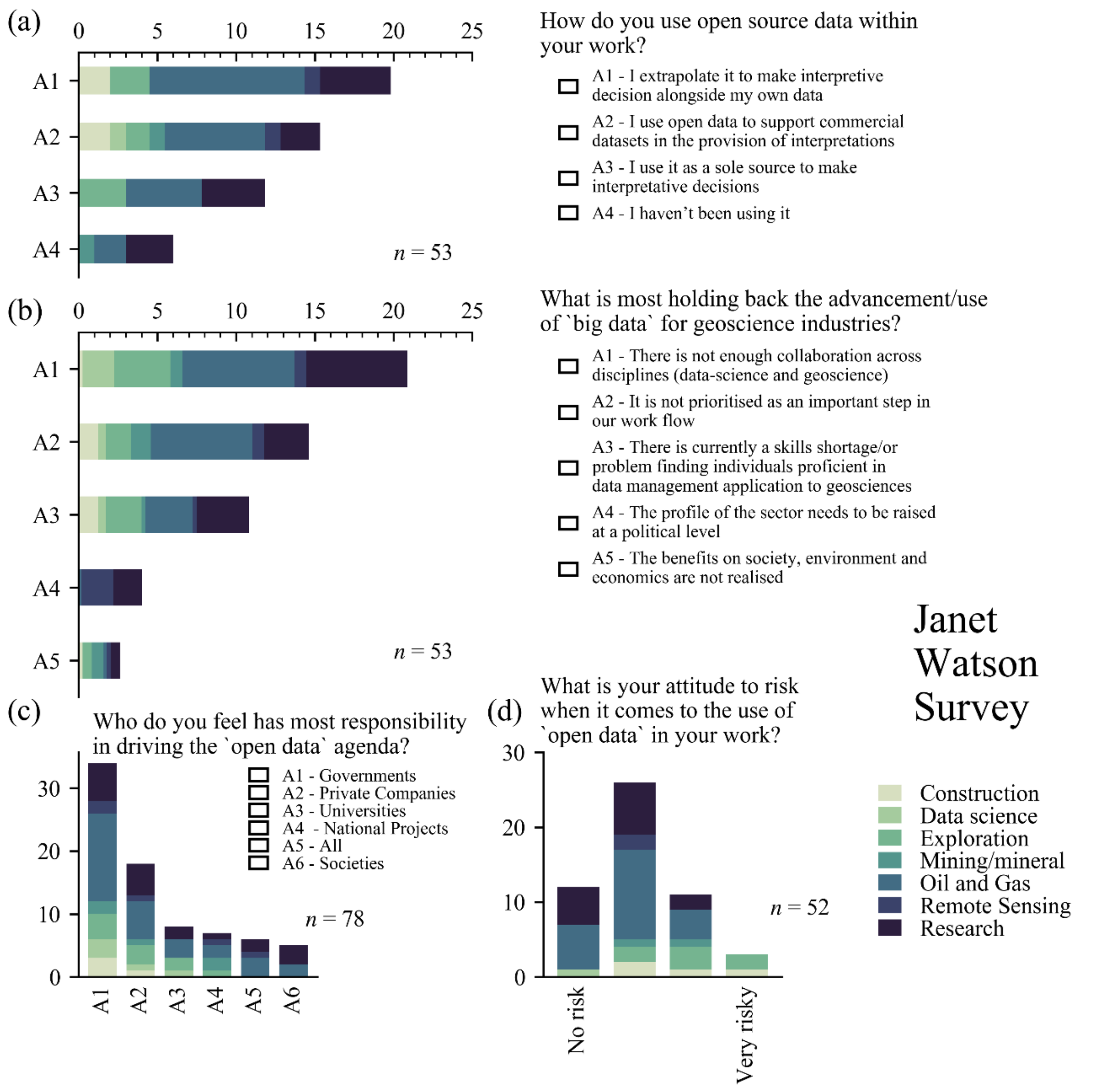

Figure 4. Results from Janet Watson Survey. The number of answers for each question is presented according to the participants sector of work (colour key) (a) shows answers to question: How do you use open source data within your work? (Q4) (b) In your opinion what is most holding back the advancements in `big data` for geoscience industries? (Q8) (c) Who do you feel has most responsibility in driving the `open data` agenda? (Q11) (d) What is your attitude to risk when using open data (Q10).

DIG TO SHARE INTERVIEW QUOTES

<table>
<tr><th></th><th>Feasibility/land procurement</th><th>Engineering Design</th><th>Construction</th><th>Maintenance and Operation</th></tr>
<tr><td>VALUE</td><td>▪ Good broad overview for tender phase, would not like to rely too heavily for detailed design, R-1.</td><td>▪ There is an advantage of having the data for scoping and to start advising designers, R-2.
▪ Boreholes can be targeted more effectively, refinement from early stages of the project, R-2.</td><td>▪ Often use BGS alongside own database, R-3.</td><td>▪ The BGS database is useful and holds vital data, R-6.
▪ Increases certainty in early stages of design, R-6.
▪ Projects would cost more without BGS, as it would slow projects down if data had to be requested, R-6.</td></tr>
<tr><td>RISK</td><td colspan="2">▪ There is a disparity in terms of quality and resolution of datasets providing challenges when doing 3D modelling, R-1.
▪ Archived information needs to be checked prior to it being shared, R-1.
▪ Open data does not take the place of the need to do site specific GI, R-1.
▪ As techniques and standards have advanced can the information be relied upon in the same way, R-1.</td><td colspan="2">▪ The risk of using acquired data is large, a stage of verification is needed, R-2.
▪ There is a risk when accepting data for which you do not know the audit process, R-2.
▪ Sharing AGS data will usually mean sharing more information than that usually held in a printed log and this needs to be checked, R-2.
▪ Liability is a big issue, someone sharing data doesn't want to be liable and equally those using the data do not want liability, R-2.</td></tr>
<tr><td>PARTICIPATION</td><td colspan="3">▪ Sharing with BGS is not at the front of everyone's minds, and is something that gets forgotten, R-2.
▪ Data sharing is non-essential so is not something to stay late at work for, R-2.
▪ There is marginal reward, clients cannot be billed for the time spent, R-2.
▪ Confidentiality of the data prevents sharing, the data can be used to inform understanding, but not legally relied upon, R-2.
▪ Public data is already being uploaded it is just private company data that needs to be pushed, R-2.
▪ It is a logistical burden, R-2.
▪ It is one of the selling points of the company that we have borehole data from lots of projects, R-2.</td><td>▪ It is not practical for commercial organisations to digitise old data, R5.</td></tr>
<tr><td>COMMUNICATION</td><td>▪ If data is made fully digital onsite there is the potential to miss steps in QA, R-1.</td><td>▪ Information needs to be transferred in the right format, R-2.
▪ There is not time or priority given to approaching project managers to describe the value of the database and sharing importance, R-2.
▪ PDF copies of GI data are being shared, R-2.</td><td>▪ Not all people in the supply chain are directly in contact with the client or are one or two levels detached from the individuals who procured the GI, R-3.</td><td rowspan="2">▪ Clients who own the data do not have an awareness of the value of the BGS database, R-4.
▪ Data sharing rules are already in contracts within government transportation projects, R-4.
▪ It is hard to find someone who knows where to insert the data sharing wording in a contract, R-4</td></tr>
<tr><td>FUTURE</td><td>▪ Its going to need an act of parliament, so it is the law to provide all borehole records within five years, R-1.</td><td colspan="2">▪ Get the government onboard and get data sharing into policy, R-2.
▪ BIM would be a good way of improving the existing system, R-2.
▪ Blanket permission to sharing data should be built into future contracts, R-2.</td></tr>
</table>

R-1 – GI contractor
R-2 - Multi-disciplinary consultancy
R-3 - Principal contractor
R-4 - Public sector body
R-5 - Infrastructure provider
R-6 - Utilities provider

Figure 5. Direct or partial quotes abstracted from original document Dig to Share (2018b), collated and sorted by interviewee role and position on a typical project lifecycle for the purposes of this research.

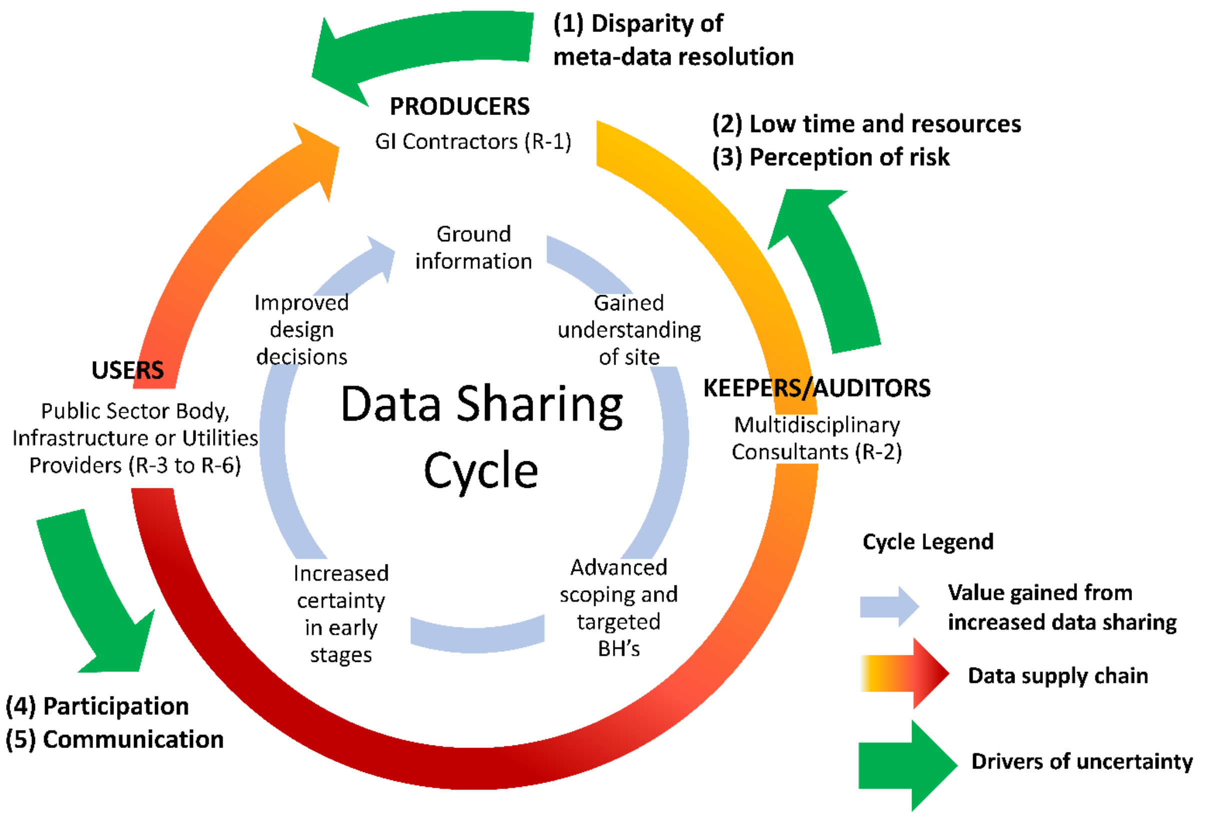

Figure 6. Summary of drivers of risk to a civil engineering project lifecycle.