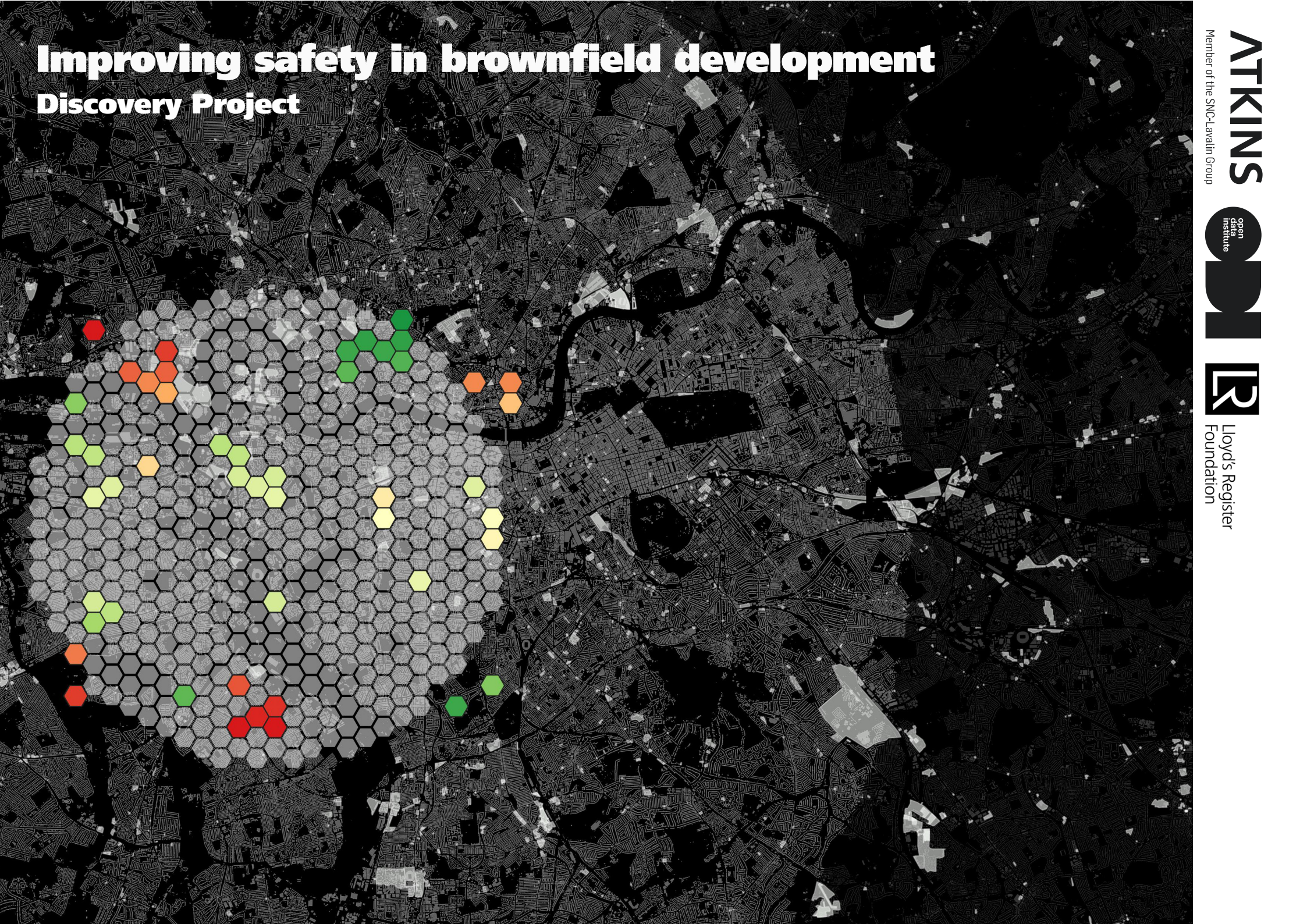


Improving safety in brownfield development

Discovery Project



ATKINS

Member of the SNC-Lavalin Group



Lloyd's Register
Foundation

Contents

Foreword	4	5 Potential Framework	31
1 Context	5	5.1 Introduction	32
1.1 Context	6	5.2 Education and collaboration	32
1.2 Below-ground data gap	6	5.3 AGS data	33
2 Our Discovery Approach	9	5.4 Open data	34
2.1 Methodology	10	5.5 Aggregated datasets	35
2.2 Key definitions	11	5.6 Framework summary	36
2.3 Audience	11	6 Summary & Next Steps	39
3 Scoping	13	6.1 Summary	40
3.1 Scoping	14	6.2 Key next steps	42
3.2 Assessing land contamination	14	6.3 Final thoughts	43
3.3 Can we capture the most important data	16	7 Appendices	45
3.4 Creating a data inventory	16	A What is a Brownfield Site	
3.5 Could AGS data be key?	17	B Land Contamination Risk Management	
3.6 Scoping summary	17	C OpenGround Cloud	
4 Case Studies	19	D Dig to Share	
4.1 Development of case studies	20		
4.2 Site-specific case study	20		
4.3 National case study	24		
4.4 Summary	29		

Version	Issue Status	Originated	Checked	Authorised	Date
01	For Information	Petra Lincoln Paul Goodship Darren Lincoln	Matt Harrison	Caroline Paradise	23/09/2020

Foreword

In 2019, the Open Data Institute and the Lloyd's Register Foundation engaged with a range of engineering and safety organisations across the UK to explore how increasing access to data can help inform engineering design, monitor safety and improve key infrastructure operations. In our report, 'Insight report on sharing engineering data', we outlined a vision for the future, potential value of sharing data across engineering sectors and systems, the barriers to sharing and the lessons we can learn from others¹.

We also created a manifesto for the engineering sector which recognises the need for leadership from across the sector, and sets out recommendations for governments, regulators, industry bodies and the private sector. This work undertaken by Atkins on data about brownfield sites drives forward a number of the manifesto points.

In our manifesto, we recommended taking a challenge-led approach to using data to address social, economic and environmental challenges. By defining a clear purpose – to support contamination assessments – and a set of common uses of data, Atkins have been able to inventory relevant data and explore how increasing access to that data could benefit developers of all sizes and increase the safety of construction.

The manifesto also recommended that data should be treated as infrastructure, be stewarded in ways that make it discoverable and accessible, and be opened and shared to unlock value. This project explores the data required to inform decision making, reduce costs and deliver on the government's commitments to develop more housing. It clearly highlights a need to make existing geospatial data about brownfield sites more open, and demonstrates how underused data held by developers could be reused in new ways.

The manifesto also highlighted the need to build data literacy and skills across professions in the engineering, construction and built environment space. It's great to see how this element is explored in Atkins outline for building multi-disciplinary teams to take forward this project.

We hope that Atkins are able to work with the Geospatial Commission and other organisations in the planning, built environment and construction sector to further develop this work. The initial findings demonstrate a great opportunity to explore new approaches to sharing that will help maximise the value of data to the industry and deliver on the UK's Geospatial Strategy. We look forward to being part of that conversation and activity.

*Jeni Tennison
Vice President and Chief Strategy Adviser at the Open Data Institute*

¹ (2019), Open Data Institute & Lloyd's Register Foundation, 'Insight report on sharing engineering data: using data for the public good', <https://www.lrfoundation.org.uk/en/news/insight-report-on-data/>

Context



1.1 Context

There are currently over 1.1 million households in England on the social housing waiting list¹, however, there is enough registered brownfield land in England to build an estimated 1 million homes². This demonstrates that brownfield land has a huge potential benefit for redevelopment, in helping to address the housing shortfall. Brownfield sites are often located in central or well-connected parts of existing town or city developments therefore presenting a prime location for housing. These types of brownfield site often provide connections to existing social infrastructure such as schools, GP surgeries and town centres. Prioritising redevelopment of these brownfield sites will also help protect greenfield land.

One of the biggest challenges in developing these sites, particularly the smaller sites with less value, is unknown ground conditions and associated risks to receptors such as human health, controlled waters, ecology and property. Unknown risks and the potential associated remediation costs make these brownfield sites less appealing to developers, particular in situations where sites are small in size and therefore offer less longer-term value to develop. Yet, there are potentially valuable sources of data that could aid in the redevelopment of these sites and encourage more development on these small to medium size brownfield sites ideally placed for housing.

1 <https://www.gov.uk/government/statistical-data-sets/live-tables-on-rents-lettings-and-tenancies>
 2 <https://www.housing.org.uk/globalassets/files/resource-files/20191030briefingmappingbrownfieldsitesinengland.pdf>

This discovery research project, funded by the Lloyd’s Register Foundation and supported by Open Data Institute via the stimulus fund³, highlights many of the major challenges preventing access to good data and discusses the industry-wide value of sharing data. This starts to address the strategic knowledge gap associated with nationwide brownfield development with the long-term ambition of creating an industry-wide framework to encourage sharing of data, facilitating a more efficient programme of house building on economically and socially suitable brownfield sites.

1.2 Below-ground data gap

There are currently concerted efforts to better understand the location of below-ground services, such as underground electricity, gas and water supplies, to minimise risk and reduce cost during construction. For example, the Geospatial Commission recently set up the National Underground Assets Registry (NUAR) to share locations of all recorded pipes and cables below ground⁴. This builds upon past projects such as ‘The Iceberg’ by Ordnance Survey and Connected Places Catapult (Figure 1), which attempted to ‘systematically develop approaches and data models to deal with the hidden infrastructure below cities’⁵. These initiatives have provided the basis for the Geospatial Commission’s five-year strategy for unlocking geospatial data to harness the potential of

3 <https://theodi.org/article/five-projects-awarded-funding-and-support-to-improve-safety/>
 4 <https://geospatialcommission.blog.gov.uk/2019/12/18/getting-under-the-surface-of-our-national-underground-assets-register-nuar-team/>
 5 <https://www.ordnancesurvey.co.uk/business-government/innovation/underground-infrastructure>

location data, with a very clear mission to ‘improve access to better location data.’⁶ This work has set a precedence for how below-ground data can be stored, managed and accessed.

It is recognised that the ground conditions within brownfield sites can have a big impact on the health and safety management of a site, site development costs and design constraints. However, the potential for contamination within the ground is not yet being recorded in the same systematic way with potential associated consequences for health and safety.

For instance, a development site can be located on or adjacent to historically in-filled land (essentially an unregulated landfill). In this scenario, there is the potential risk that ground gas can migrate to a development and result in associated health issues for the occupants. Also, in a recent example a residential development that had been unknowingly built on a burning coal seam had to be demolished due to the detection of high levels of poisonous carbon monoxide⁷.

These risks pose less of a problem for large, experienced developers who can afford to purchase the biggest sites, where the economy of scale is such that the necessary investigations can be commissioned, and the associated health & safety risks mitigated. Smaller, less experienced, developers

6 Unlocking the power of location: The UK’s geospatial strategy (2020) - <https://www.gov.uk/government/publications/unlocking-the-power-of-locationthe-uks-geospatial-strategy>
 7 Example from <https://www.gov.uk/government/publications/unlocking-the-power-of-locationthe-uks-geospatial-strategy>

who potentially cannot afford this initial investment, are less likely to take the risk on purchasing and developing these brownfield sites. This may have the consequence of leaving many well-positioned but smaller, inner-city brownfield sites undeveloped and may encourage smaller developers to construct on safer greenfield sites, often on the outskirts of the city. From all the 24,547 brownfield sites currently registered by English local authorities, the median number of potential new dwellings a site could provide is 12. This demonstrates that the average brownfield site is best suited to a small development, yet as discussed, these are often present more risk, either to health and safety or financially for the developer.⁸

Our research looked at data that could help identify potential site development constraints (or abnormal costs) associated with a site’s brownfield status and previous uses. This work hopes to support early identification of potential risks during site selection which should aid developers to pursue these types of sites. In this report, the project team demonstrates that significant additional value and benefit could be unlocked if data is shared more freely. Demonstrating how datasets could potentially be aggregated for initial high-level ‘screening’ of sites for potential development. The team has broken down the challenge into three challenge statements shown to the right.

8 Data from Ministry of Housing, Communities and Local Government (MHCLG) - <https://digital-land.github.io/dataset/brownfield-land/>

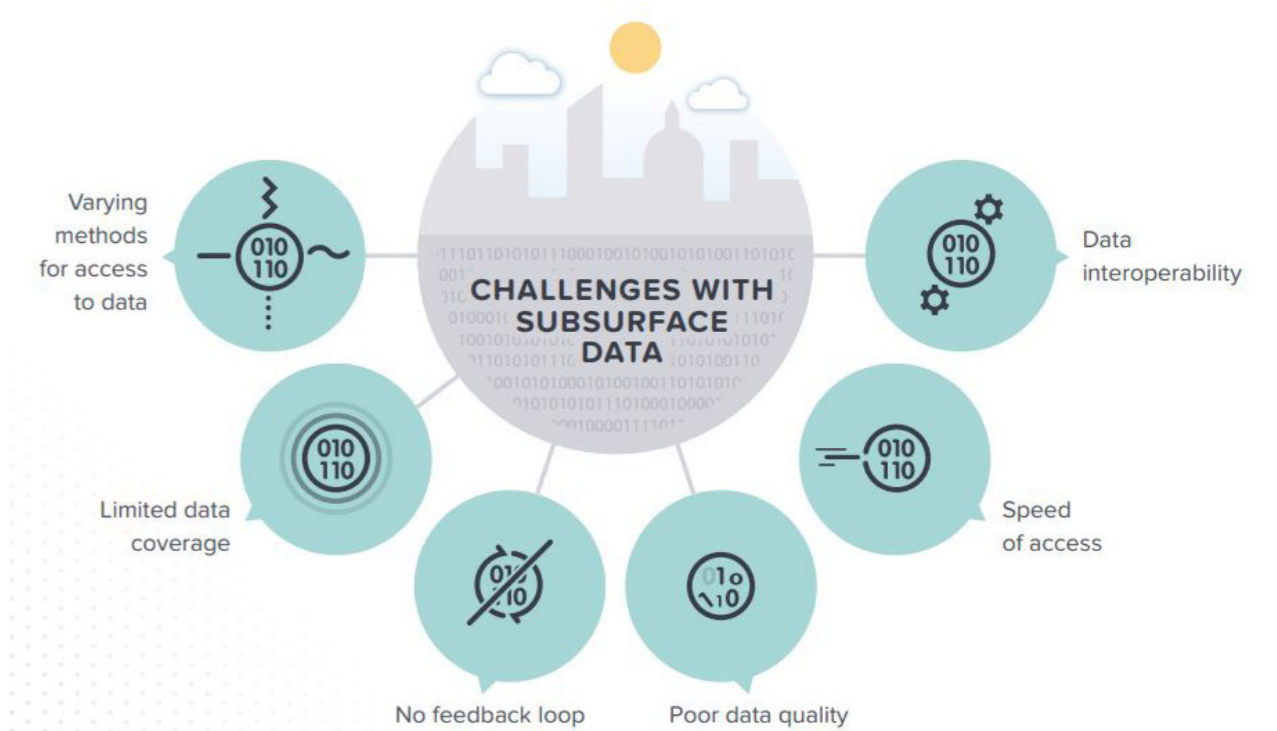


Figure 1 - The problems associated with subsurface data. Project: The Iceberg by Ordnance Survey and Connected Places Catapult.

Challenge Statements

1. Accurately catalogue data that could provide lines of evidence to the levels of contamination on brownfield land.
2. Analyse case studies to demonstrate how open and closed data for potentially contaminated land could greatly reduce unknown risks on site.
3. Create a framework for encouraging data sharing across the contaminated land industry.

Our Discovery Approach

2

2 Our Discovery Approach

2.1 Methodology

To answer these challenges, this discovery project explored the viability of unlocking data used within land contamination risk assessment using a three stage approach. The first stage gathered as much information as practical, the second stage utilised the data gathered to conduct detailed case studies and finally, the third stage, used the results from the previous stages to create a framework.

The key outputs of this study are:

- This report, which formally documents our research, including objectives, methodology, findings and next steps.
- A webinar, where our research is presented to wide audience, alongside general problems associated with land contamination data.¹
- An open access Data Inventory data relating to land contamination.²
- Sharing code for accessing data from proprietary database and aggregating it at a national scale.³

¹ <https://www.youtube.com/watch?v=1V8U3CA1AIU>

² <https://github.com/atkinsglobal/ground/>

³ As above (Code not yet shared)

Stage 1: Scoping

During the first project stage, the team conducted an initial mapping exercise with the ODI, relying on the breadth of our internal networks to establish a wide spectrum of discipline knowledge. This mapping exercise was then followed by a desktop research stage to categorise the main themes and identify the important data sources. During this stage the team also developed a data inventory and shared it publically on GitHub to enable others to benefit from the knowledge gained during this stage.



Stage 2: Case Studies

During the case study stage of the project, the team completed two detailed studies at different scales, to compare open and closed data sources and determine the resolution of data available in the OpenGround® Cloud database.

The site-specific case study was conducted on a single site and investigated the difference between open and closed data. The national case study looked at the resolution of the open data and investigated the OpenGround Cloud database used by Atkins for storing project information.



Stage 3: Framework Development

Finally, drawing on the case study and scoping stage, the team developed a framework to enable better data sharing in the industry. This stage investigated the particular difficulties and case studies used in the previous stages to identify key activities and good practices that need to be established in order to standardise the use of data for assessing brownfield development.



Figure 2 - Three-stage linear exploration process

2.2 Key definitions

2.2.1 Brownfield Land

A brownfield site is generally understood as “previously developed land”⁴ that has the potential for being redeveloped. It is often (but not always) land that has been used for industrial and commercial purposes and is now derelict. In England “The Town and Country Planning (Brownfield Land Register) Regulations 2017”⁵ require local authorities to prepare and maintain registers of brownfield land that are suitable for residential development. These provide details of the minimum number of new dwellings that could be developed on each, and information on the planning status and ownership of each site. More details on brownfield registers can be found in Appendix A.

2.2.2 Contaminated Land Risk Assessment

Contaminated land generally refers to land that contains elevated concentrations of potentially hazardous substances. These concentrations may be present naturally, but more commonly, contamination is the legacy of the industrialisation of Britain over the past 200 years. Land that is contaminated can present an unacceptable risk to human health, the environment (organisms, air, soil, the subsurface, groundwater and surface water), and buildings and structures.”⁶

⁴ <http://www.sustainablebuild.co.uk/brownfieldsites.html>

⁵ <http://www.legislation.gov.uk/ukxi/2017/403/contents/made>

⁶ CIRIA C552 - Contaminated Land Risk Assessment: A Guide to Good Practice (2001)

2.3 Audience

This report is intended to provide information for people in roles across the breadth of the data and construction industry, particularly those who are working in residential development and those who are working on brownfield sites. The study focuses on the use of contaminated land data, however is not an exhaustive guide to the contaminated land risk assessment process and should not be used as a technical guide. The information provided is intended to encourage data sharing and use across the industry, whilst outlining potential issues and a framework for future development.

Key

Number of Brownfield Sites

- 0 - 9
- 9 - 16
- 16 - 23
- 23 - 28
- 28 - 34
- 34 - 41
- 41 - 51
- 51 - 72
- 72 - 108
- 108 - 482

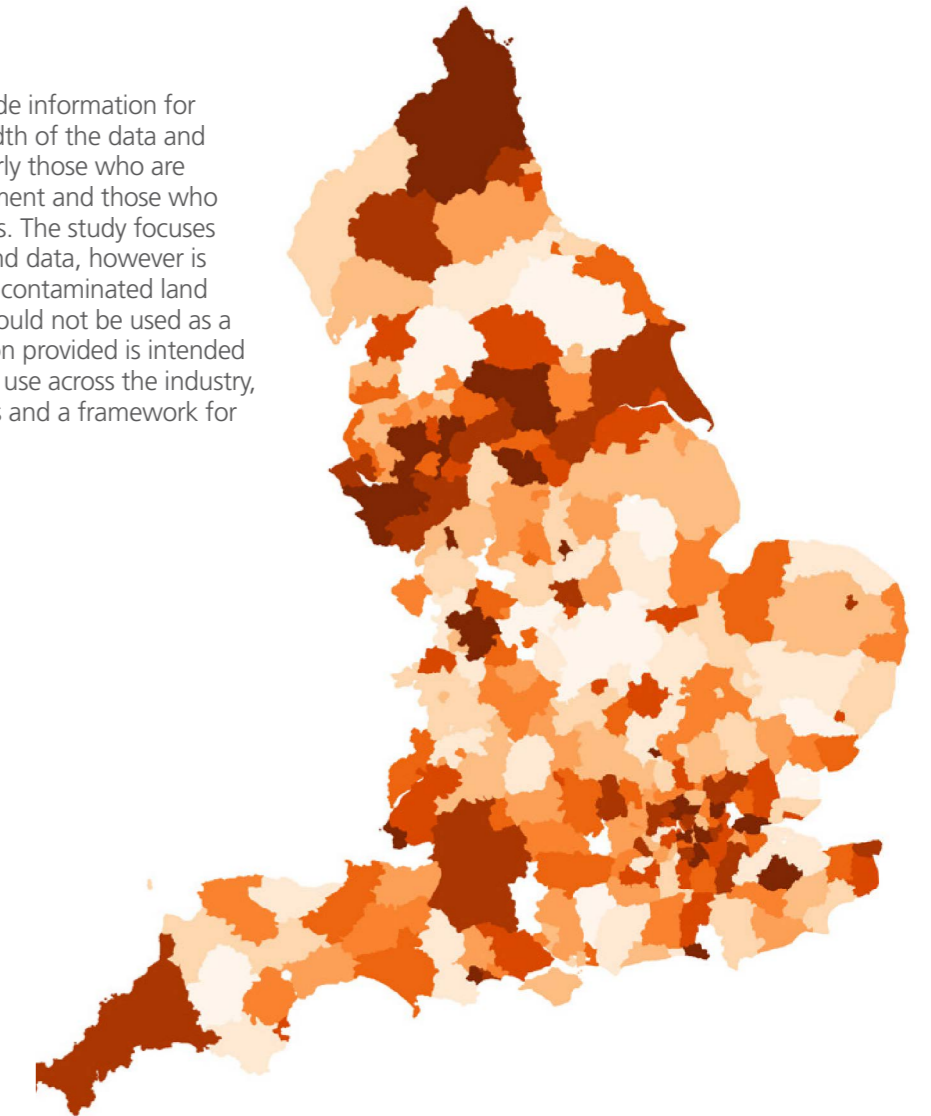


Figure 3 - Map of brownfield sites in England

Scoping 3

3 Scoping

3.1 Scoping

The early stage scoping phase focused on understanding the breadth of data available to examine the concentrations of contaminants within the ground. To do this, an initial review of current literature relating to ground and land contamination data was completed. At the same time, a coherent list of relevant data, both open and closed, was catalogued into an openly licensed data inventory for a wider community to input and build over time. This provided the team with an opportunity to create the first open access data inventory for land contamination, that could potentially help inform others examining brownfield sites.

3.2 Assessing land contamination

The assessment of land contamination is typically completed in three stages using the Land Contamination: Risk Management (LCRM) guidance¹. This starts with a Tier 1 Preliminary Risk Assessment (PRA), which typically involves a desk based review of information and a site walkover. The output of the Tier 1 PRA is a Conceptual Model (CM) for the site which provides details on the risks from site-specific contaminants identified and recommendations for further investigation. This is then followed by a Tier 2: Generic quantitative risk assessment (GQRA) that typically involves site investigation, monitoring and comparison of measured concentrations of targeted contaminants within soil, soil-derived leachate² and

¹ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

² Leachate are chemicals which can leach from the soil and pore water.

groundwater to generic assessment criteria (GAC). Should unacceptable risks be identified following this stage, a Tier 3: detailed quantitative risk assessment (DQRA) may be recommended. This involves modelling of fate and transport of contaminants and comparison of measured concentrations of chemicals with site specific assessment criteria (SSAC).

Crucial to each stage is the Conceptual Model (Figure 4). During Tier 1 risk assessment, an Outline CM identifies potential pollutant linkages via the source-pathway-receptor relationship. A risk only exists when a potential source of contamination, pathway for migration of contamination and key receptors are identified. As the guidance states "without a pollutant linkage, there is not a risk - even if a contaminant is present.". Additional data is then collected during Tier 2/Tier 3 risk assessment and the risks identified in the CM are updated at each stage.³

Remediation needs to be undertaken in the case that unacceptable risks are identified during the risk assessment.⁴

³ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

⁴ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

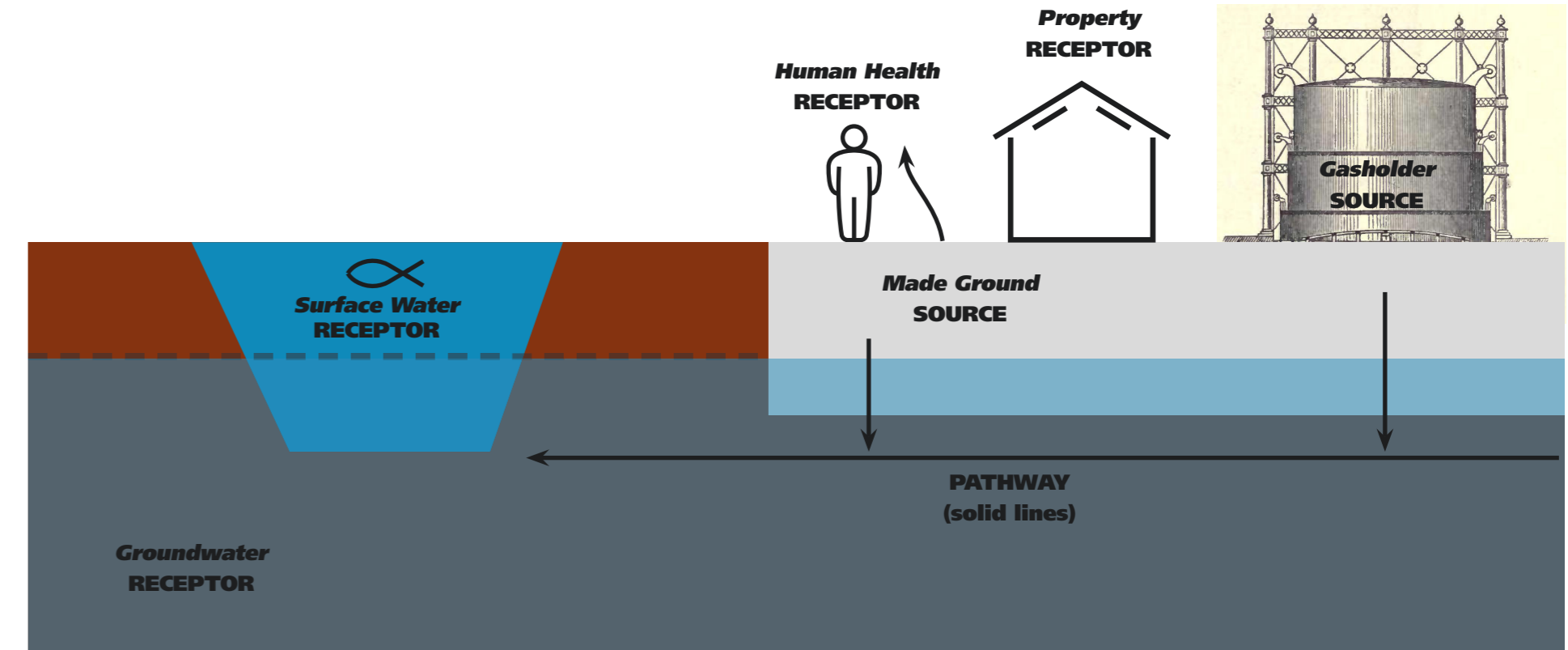


Figure 4 - A basic visualisation of a conceptual model¹. British Standards Institute (BSI) publishes BS EN ISO 21365:2020 Conceptual site models for potentially contaminated sites. This document provides guidance on developing and using conceptual models through the various phases of investigation, remediation, and any subsequent construction or engineering works. It describes what conceptual models are, what they are used for and what their constituents are.

¹ 'Conceptual model' is a term used within the latest LCRM guidance. Previous guidance (CLR11 - Model Procedures for the Management of Land Contamination 2004) and BS EN ISO 21365:2020 use the term 'conceptual site model'. 'Conceptual model' and 'conceptual site model' are interchangeable terms

3.3 Can we capture the most important data

Data used within the land contamination risk assessment is typically a combination of open data and closed data⁵. This usually comes from a variety of sources and owners, and in a wide range of file formats. For example, you might get the borehole locations in a CSV⁶ file format from the British Geological Survey (BGS) and historic land uses from Groundsure in GIS shapefiles. These data types then need merging or to be overlaid, so that they can be cross analysed, which is a very typical workflow. It should be noted that datasets licensed by Landmark or Groundsure are typically purchased for use in the PRA, thus increasing the cost of assessment.

The cataloguing of datasets provides a very valuable insight into what is currently available for a specific topic and can provide some very basic information such as file type, geographic coverage and licensing area of analysis⁷. For this research project, a data inventory has been created for the land contamination risk assessment process (from Tier 1 to 3), so that an open access list of data available could be made available to a wide audience. This could then be shared and developed by a wider community over time.

⁵ Closed data here is commercially licensed data that has to be purchased as its owned by a 3rd party company.

⁶ Comma Separated Values, a common data format.

⁷ How to create a data inventory - <https://gatesopenresearch.org/documents/2-45>

3.4 Creating a data inventory

To complete the data inventory, all the data necessary to produce a PRA desk study report (Tier 1), a GQRA report (Tier 2)⁸ and a DQRA report (Tier 3) was documented. This included both open and closed data. This list of data sources was pulled together using professional knowledge from experts who regularly compete risk assessments, however, over time it is hoped that this list will grow and become more refined with it being open access. The information included within the inventory is as follows:

- Dataset Name
- Information Included
- Format
- Pollutant Linkage (does data provide information on sources, pathway and receptors)
- Assessment Stage (Tier 1,2 or 3)
- Data Owner
- Source
- Link to data
- Date Created
- Frequency of updates
- Licence
- Confidentiality
- Automated Processing
- Notes
- Geographic Coverage
- Importance Ranking

⁸ Containing information from ground investigation.

The compiled data provides information on historical and recent land use, geology, hydrogeology and hydrology, mining, radon, landfill and waste sites, pollution controls and prevention, unexploded ordnance, risk assessment criteria (GAC and SSAC), site investigation data, etc.

Professional judgement was used to prioritise initially the importance of each dataset, however it is hoped more formal prioritisation could develop over time as more professionals collaborate on the data inventory. Statistical analysis could be used to better understand relationships and provide a basis for prioritisation of the data at each stage.⁹

⁹ The data inventory is shared here: <https://github.com/atkinsglobal/ground/>

3.5 Could AGS data be key?

Within this data inventory, the AGS data format¹⁰, as prepared by the Association of Geotechnical & Geoenvironmental Specialists (AGS)¹¹, is of particular interest for the land contamination risk assessment process. This is because it enables the management of detailed geo-environmental data from ground investigations, including geological descriptions, details of geotechnical and geo-environmental samples, and the results of laboratory analyses of chemicals.

The AGS data format provides an industry standard for structuring Association of Geotechnical & Geoenvironmental Specialists data. Data in this format is commonly referred to as AGS data. For example, it specifies that various data groups should be identified by a four letter abbreviation and columns within the groups are identified with further standard abbreviations.¹²

Many stakeholders within the ground engineering and land contamination industry manage their data via the Bentley Systems proprietary data management cloud platform, OpenGround Cloud. Provides a robust centralized data repository for traditional project-by-project orientated deliverables, with the associated Web API enabling cross project data mining and extraction. More detail can be found in Appendix C.

¹⁰ *.ags is a text file format used to transfer data reliably, between organisations in the site investigation industry, independent of software, hardware or operating system.

¹¹ <https://www.ags.org.uk/data-format/>

¹² Sample information is provided in the SAMP group. Date and time the sample was collected is recorded in SAMP_DTIM column.

This detailed data can provide specific information on the concentration of contaminants on a particular site and identify local risks. However, given that this data is generally used once for single project purposes due to traditional industry practices and systems not supporting multi-project data aggregation, this valuable data will rarely get reused for further strategic analysis.

3.6 Scoping summary

The discovery phase focused primarily on understanding the breadth of data available to land contamination risk assessment. To do this, a coherent list of relevant data was catalogued into an open access inventory, with the ambition that over time a wider community could input and build upon this, allowing the most relevant data to become more easily accessible. From this list it becomes apparent that AGS data contains information collected during ground investigation and could therefore provide site-specific information on ground conditions. This initial scoping exercise shows, that if AGS data together with other closed data were readily available, our understanding of the ground conditions would significantly improve.



Case Studies

4

4.1 Development of case studies

The second challenge statement required the team to develop case studies to demonstrate how open and closed data of previously developed sites could reduce uncertainty with respect to ground condition. Following the creation of the data inventory in the initial stages of the project, the team determined that two main case studies should be undertaken as part of the research.

Firstly, a site-specific case study was undertaken to assess the difference between the outcomes of the risk assessment process when using only open data versus using closed data. The chosen site was anonymised due to commercial sensitivities. Then, a national case study investigated the possibility of assessing risk on a national scale by comparing concentrations of selected chemicals to GAC. This study aimed to demonstrate the gaps in open data and highlight the benefits of making data available such that the risks associated with ground contamination can be better understood.

4.2 Site-specific case study

The site-specific case study took the form of a preliminary risk assessment (PRA – Tier 1) for an unnamed former gasworks site that is currently being developed for residential end-use. The risk assessment was carried out in accordance with the Land Contamination: Risk Management Guidance¹.

During the PRA stage of an assessment, information is collected via a desk study and a site walkover. Based on this information, a conceptual model is developed to identify potential pollutant linkages and associated risks². It should be noted that for this research no site walkover was undertaken as part of this project. This is standard practice, where practicable, for a PRA. The typical process is shown in Figure 5.

¹ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>
² <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

Gas Sites

Gas was manufactured in Britain between 1792 and 1981 and is typically produced from coal, generating a variety of waste and by-products such as coal tar, ammonia, phenol, cyanide and sulphate. As a result, contamination by these by-products can be typically identified at former gasworks sites¹. These historical landmarks can often provide clues to potential contamination. The site chosen for this site-specific case study also has a nearby gasworks providing some initial clues to potential contamination.

¹ CL:AIRE: Gasworks Profile A – The History and Operation of Gasworks (Manufactured Gas Plants) in Britain

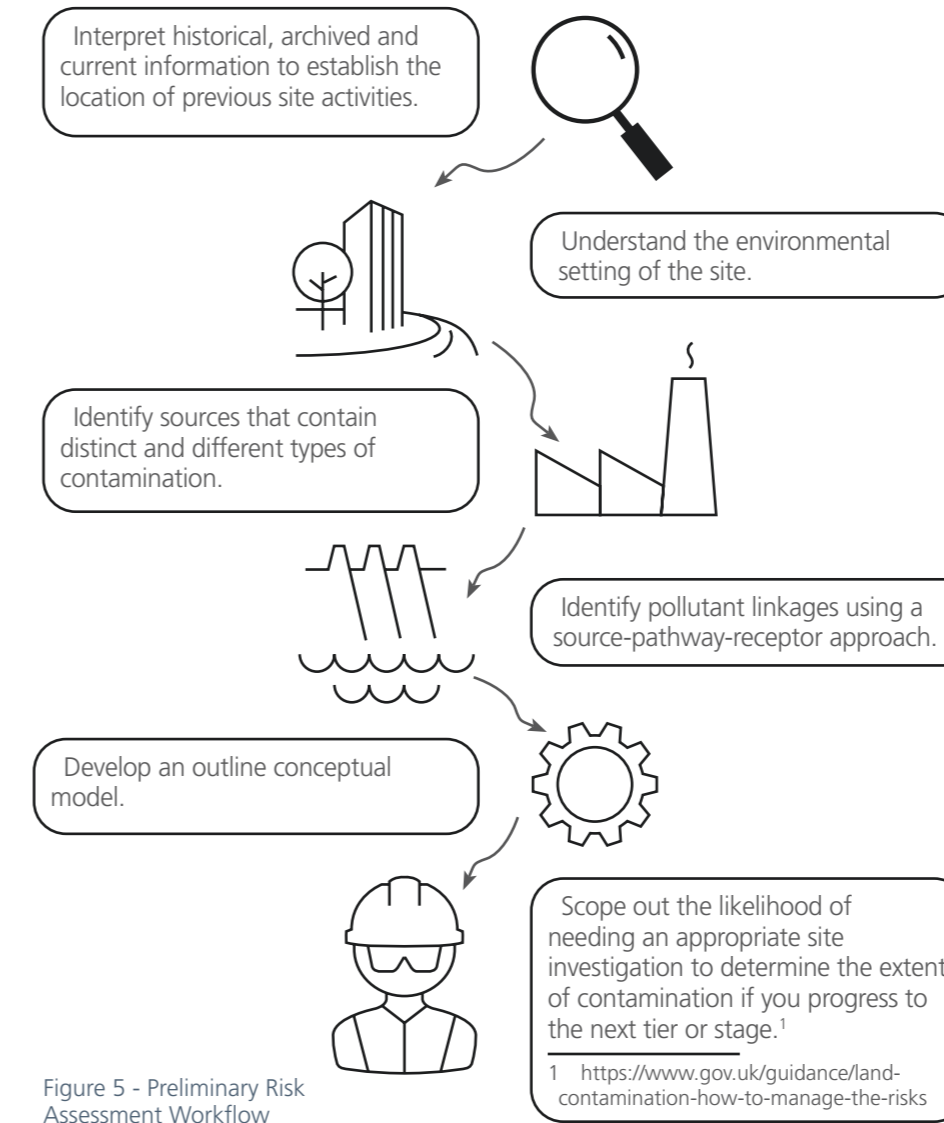


Figure 5 - Preliminary Risk Assessment Workflow

4.2.1 Site-specific case study data

For the site-specific case study Groundsure³ geospatial data for the site as well as a 500 m buffer area was purchased. This provided the opportunity for two different conceptual models to be created, one using only open data and the other using closed data as well as open data. Furthermore, this approach allowed for a comparison between information available within the open data and closed data to be undertaken for the same site. The datapoints used in the PRA are shown in Table 1. The closed dataset included further information on all datapoints openly available, with the exception of unexploded bombs.

³ Groundsure are a 3rd party company who sell/license environmental setting data on sites.

Datapoint	Status
Current Land Use	Closed only
Contaminated Land Register Entries	Open
Discharge Consents	Open
Geology & Coal Mining	Open
Historical Land Use	Closed only
Hydrology	Open
Hydrogeology	Open
Landfill Sites	Open
Other Mining (not coal)	Closed only
Pollution Incidents	Closed only
Radon	Open
Sensitive Land Use	Open
Unexploded Bombs	Open
Waste Sites	Closed only

Table 1 - Open versus closed data points

4 Case Studies

4.2.2 Site-specific case study findings

The open data provided information on the environmental setting of the site. However, information on site-specific sources and historical site activities and any associated contamination was very limited. Furthermore, open data identified artificial ground and two tank like structures in the northern part of the site⁴. A BGS geological log, which shows information from a historical borehole, was available for the central part of the site. This indicated that a well was drilled to provide a water supply for the gasworks.

The conceptual model which was created with the open data provided very limited evidence of the site being previously used as a gasworks site. To obtain information on the historical onsite gasworks, historical mapping and historical industrial land use data was needed. However, as previously outlined, this type of data and mapping is rarely available from open data repositories. As a result, an open data conceptual model could omit important historical/current sources of contamination used to inform the PRA.

It is also worth noting that additional information about the site was available on the local authority's planning portal. This included conceptual models, data from previous ground investigation reports, and an Envirocheck report as part of the environmental statement for the site. It was assumed as part of the environmental statement that remediation would need to be undertaken for the site. Furthermore, several potential data sources, such as the geological

logs, are also included on the planning portal, but these are not in a machine readable format. Nor is there any clear licensing information about the reuse of this data. This indicates a huge potential source of valuable, publicly available, historic information which could potentially be reused if published under a licence.

Purchasing closed datasets from Groundsure or Landmark makes land contamination risk assessment processing easier as it fills in most of the gaps in the data mentioned above and brings it all together in one place. From these datasets we can compile information on historical land-uses, which is vital to understanding potential contamination. This data, alongside data such as artificial land extents, allows us to develop a more accurate conceptual model and risk ratings for receptors during the PRA.

⁴ BGS Geoindex & OS Open Data

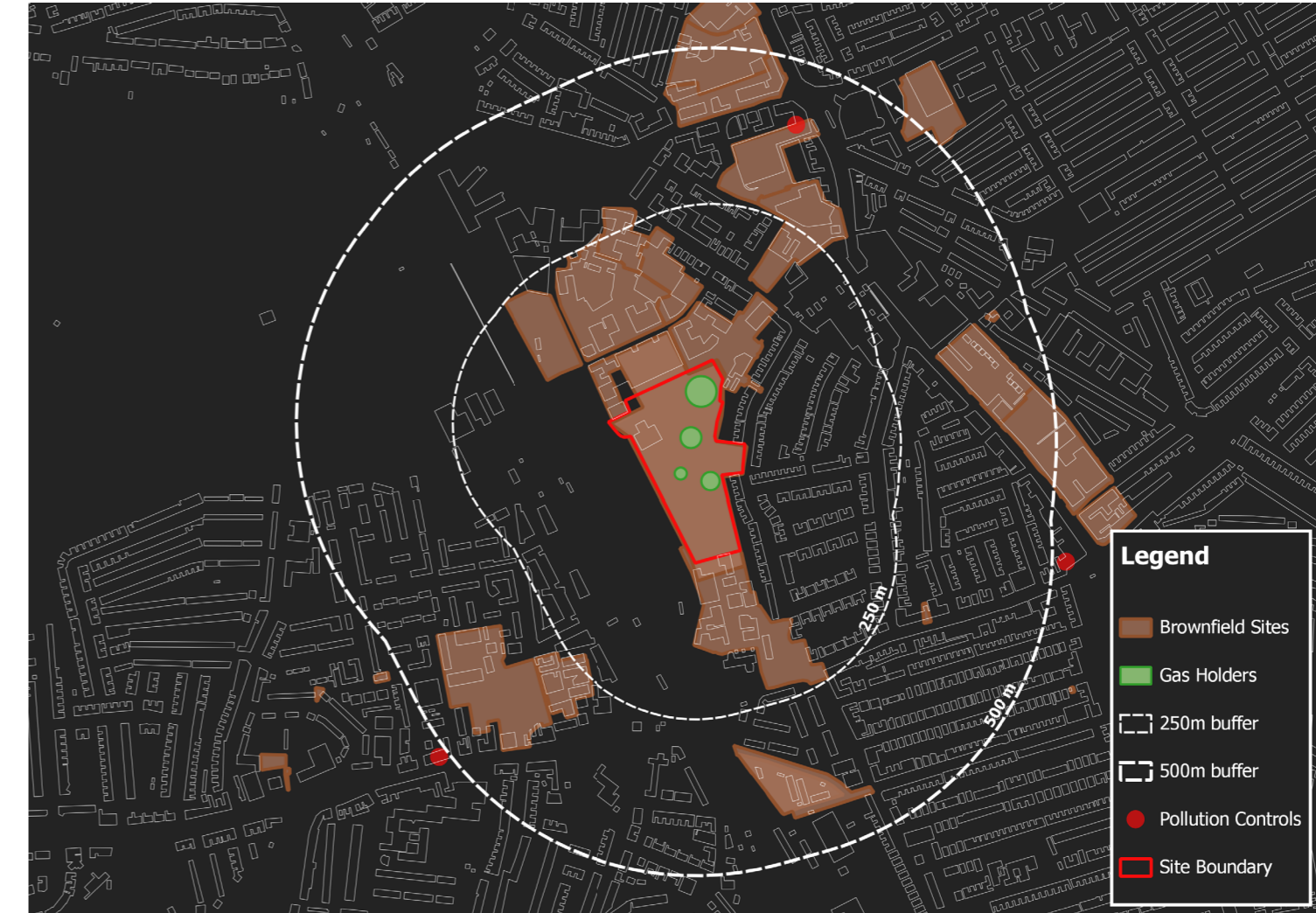


Figure 6 - Case Study Site

4.3 National case study

From work previously completed by Atkins⁵, we can see in England alone there are 24,547 brownfields sites across England, these offer the potential of approximately 1,250,000 new dwellings as a minimum. This clearly has the potential to significantly impact the housing crisis in England. At a more local scale we can also see the potential impact, as Haringey London Borough Council for example, a typical inner-city local authority, data indicates that there are 244 brownfield sites with up to 20,974 potential new homes covering 182 hectares of land⁶. The ten year housing target for the borough is 15,019 new homes⁷. This shows that there is enough brownfield land within the borough of Haringey to provide land to achieve the ten year housing target .

However, simply finding a location available to develop is not enough, there needs to be a clear understanding of the risk the land poses for it to be taken forward for redevelopment.

The data inventory shows us that site-specific geospatial data can be obtained to aid in our understanding of the conditions on identified brownfield areas. For example, 123 of 244 brownfield sites are located within a groundwater Source Protection Zone (SPZ) (Figure 7) and 68 sites are underlain by superficial deposits. However, the open

⁵ Edaroth White Paper - <https://edaroth.co.uk/wp-content/uploads/2020/01/EDAROTH-Whitepaper-January-2020.pdf>
⁶ MHLCG - <https://digital-land.github.io/dataset/brownfield-land/>
⁷ <https://www.london.gov.uk/what-we-do/planning/london-plan/current-london-plan/london-plan-chapter-three-londons-people/policy>

data available provides only general information on the environmental setting of the sites at a large-scale resolution as demonstrated by the availability of SPZ and superficial geology data.

As highlighted by the site-specific case study, crucial information needed to underpin the land contamination risk assessment for a series of sites at a strategic scale is not available from open sources.

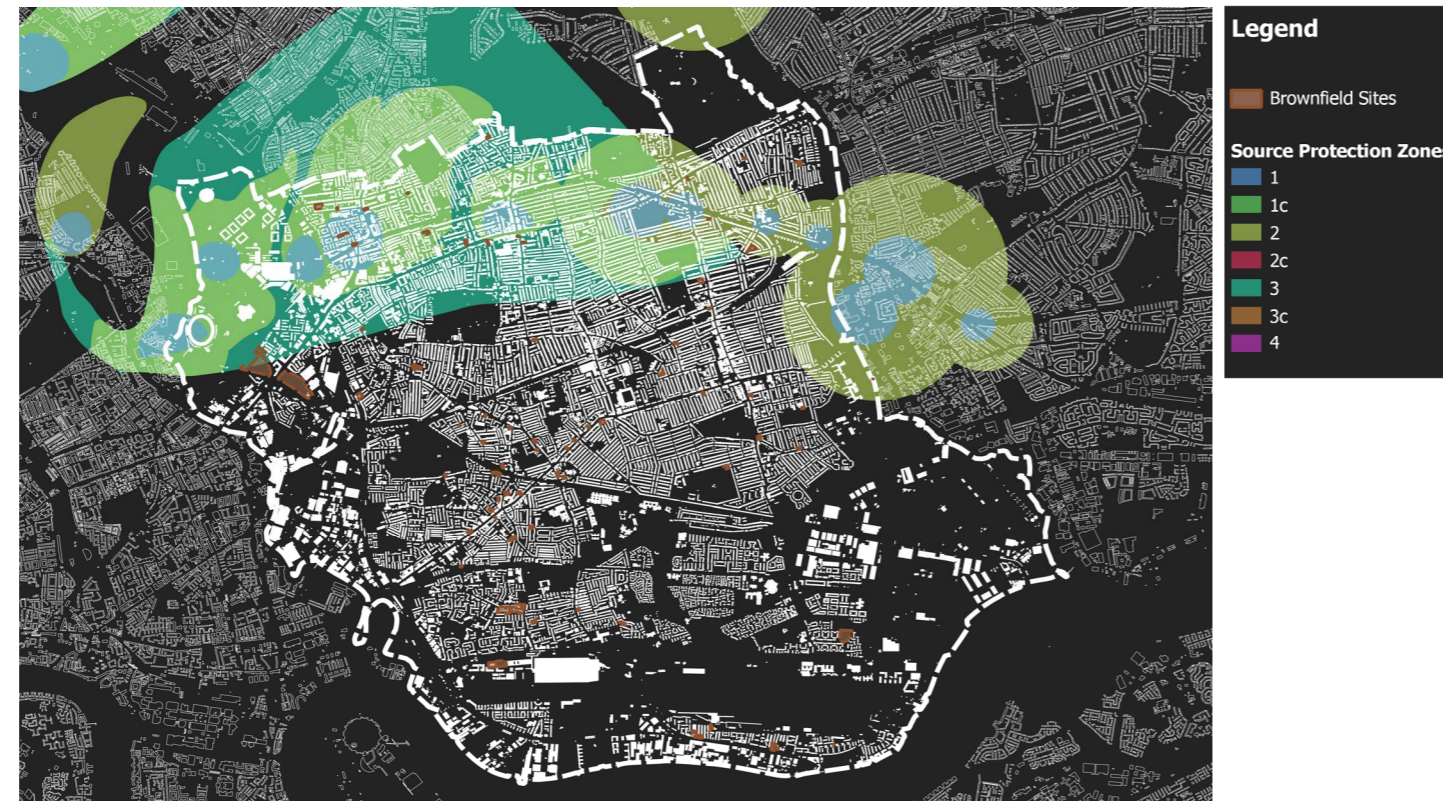


Figure 7 - Source Protection Zones

4.3.1 Big Data: Is there a better solution using AGS data?

As well as reviewing the open data, the national case study included a review of the data available to Atkins via the OpenGround Cloud database. The OpenGround Cloud database is used by Atkins for collecting and storing ground investigation and monitoring data on behalf of its clients. This database contains a large AGS dataset for many projects (traditionally accessed on a project-by-project basis). Similar systems are used by other consultants using the same platform, but importantly these databases are not interconnected.

For this research the data from each project in the OpenGround Cloud database was extracted via its web API⁸. All the available multi-project data was then merged for analysis, creating a unique individual data-frame of all the project data. This on its own could potentially be extracted as a single file, such as a CSV, to analyse the historic location of boreholes or to create a dashboard for project to interrogate past projects. The future ambition is to integrate this data with other geospatial datasets for ongoing aggregation, scaling and analyses. This would have the potential to allow other geospatial datasets from the same location to be joined, allowing for a more accurate digital representation of the built environment to be created. For this discovery project, the ground investigation data was reused and aggregated at a national scale within Atkins in order to understand if this compiled dataset could potentially reveal insights into contaminant patterns for brownfield sites.

⁸ Application Programming Interface

For this extraction the data analysed included:

- Exploratory hole location details.⁹
- Types of samples and depth at which they were collected from the exploratory hole or well.¹⁰
- Results of chemical analysis for the samples.¹¹

It should also be noted that there is significantly more information related to land contamination risk assessment within the OpenGround Cloud database including encountered geological strata, groundwater and gas monitoring, headspace testing using a photo-ionisation detector, variable head testing, etc. However, because this research is focused on discovery and not a detailed analysis of one specific dataset, these datasets were not reviewed as part of this project. Other datasets could also potentially offer great value and should be fully explored in a similar manner in future research.

⁹ Location Details – LOCA group within AGS
¹⁰ Sample Information – SAMP group within AGS
¹¹ Environmental Contaminant Testing – ERES group within AGS

4.3.2 Screening the data

In order to provide an indication of the concentrations of contamination across the multi-project/national dataset, two representative contaminants were selected from this large merged dataset:

- Benzene – a typical representative of highly mobile (within water) and volatile organic compounds.
- Benzo(a)pyrene – a compound typically used to assess risk posed by a range of polycyclic aromatic hydrocarbons.

To select the relevant data, the chemical name column within the dataset was filtered for “benz” and “Benz”. Benzene and benzo(a)pyrene data entries required further bespoke filtering, due to the inconsistency in AGS chemical naming convention contained within the column, even for the same chemical compound. Typical laboratory testing results are reported for soil, soil-derived leachate, waste acceptance criteria (WAC) testing and water samples. Differentiating between these types of tests within AGS is problematic due to inconsistencies in data reporting between different laboratories¹². However, some rules are applicable to all data:

- All groundwater and surface water samples are defined as Type (SAMP_TYPE) “EW” or “W”.
- All other samples (soil, leachate & WAC) are generally covered under the “ES” or “COMP” Type. Differentiating between them depends on each lab’s procedures for producing AGS data.¹³

¹² Typical laboratory could be for soil, soil-derived leachate testing, waste acceptance criteria (WAC) testing and water testing.

¹³ For more detail on the meaning of abbreviations used within AGS data, see AGS data format website. <https://www.ags.org.uk/data-format/>

In light of the AGS data inconsistencies, data was processed using the approach shown in Figure 8. The soil results represented by readings in mg/kg and/or ug/kg were compared to the strictest generic assessment criteria (GAC) for assessing the risk to human health receptors. The soil-derived leachate and water results were compared against the strictest GAC for assessing the risk to controlled waters receptors.

However, this approach has the following limitations:

- Composite samples that have not been recorded as “COMP” Type and are labelled as “ES” were included in the analysis;
- Soil and leachate results for tests with units other than mg/kg, mg/l, etc. such as pH were omitted from the analysis;
- Some of the WAC testing reported with mg/kg units were falsely included in the soil analysis; and
- Leachate preparation techniques were not differentiated. Hence, various leachate preparation techniques were treated in the same way, included those of WAC testing.

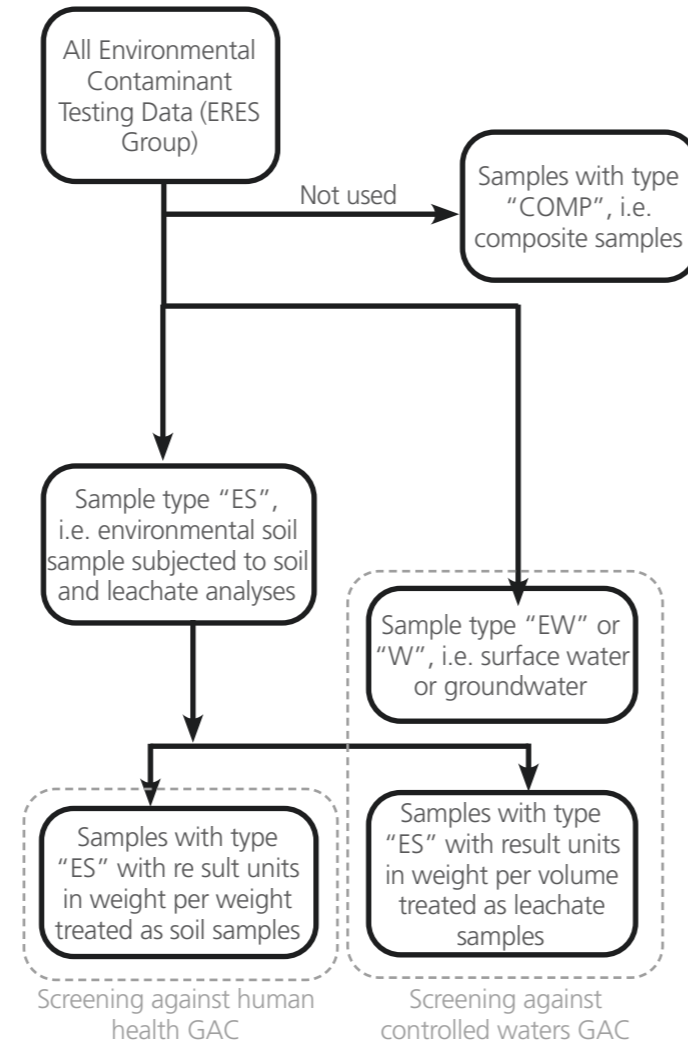


Figure 8 - AGS data processing approach

4.3.3 Visualisation and further analysis

A new dataset was then created using the rigorously screened data. This dataset could not be confidently analysed and this provided an initial opportunity to spatially explore and analyse the results within a geospatial platform. In order that results could be visually conceptualised, a simple scoring system was used. Where an exceedance of GAC was noted indicating potential risk, a value of 1 was assigned to the sample. Where no GAC exceedance was noted, -1 was assigned to the sample. This allowed a numerical value between -1 and 1 to be calculated for any given group of boreholes by calculating the average score for the area. This could then be visualised within a powerful web-based tool like Kepler¹⁴, where the numerical value would be coloured appropriately, potentially mapping all boreholes in the UK. This allowed a hexgrid to be generated from the boreholes extracted from OpenGround Cloud database using hexagons with a 250 m diameter. (Figure 9).

The hexagonal representation of the screening is a proof-of-concept of visualisation of screening results. Screening could be done for additional contaminants of concern and data could be integrated with other datasets listed in the data inventory, including other ground investigation data available in AGS format. The final hexagonal dataset could provide information on the estimated risk posed by land contamination prior to undertaking PRA. This is discussed further in the Potential Framework section.

¹⁴ WebGL - <https://kepler.gl/>



Figure 9 - Conceptual visualisation idea for national scale benzene and benzo(a)pyrene screening against GAC. Green denotes no GAC exceedances, with dark red indicating all locations within the hexagonal area having GAC exceedances.

4 Case Studies

4.3.4 Encountered problems during case studies

While the AGS data format potentially provides great values, it should be noted that the data within this format is generated by environmental specialists for their specific use. The dataset must therefore be interpreted by an expert and in the context of other datasets. As a result, it is not possible for any general conclusions to be drawn from the dataset on its own regarding land contamination risk. It is also worth noting the data collection on site is often completed in very traditional ways, such as via handwritten notes that are later inputted into a spreadsheet/database, nor is the inputting of the data always very consistent. This results in domain specialists being required to quickly and easily interpret these certain nuances in the data, preventing others from easily interpreting the data.

Some online datasets, such as aerial maps, do not have clearly defined terms and conditions for use within a commercial setting. This uncertainty often prevents people from using the data or potentially results in the inadvertent reuse of third party data without permission.



Figure 10 - Example ground investigation

4.4 Summary

Data currently published as open data does not enable land contamination risk assessments to be undertaken in accordance with LCRM guidance whilst satisfying the technical requirements of the process. This is because critical sources of contamination could be missed if only open data is used for the assessment.

AGS data could aid the industry to make initial assessments of risk if they currently store large amounts of this data locally. This is discussed further in the Potential Framework section of this report. The AGS data could enable these initial risk assessments by allowing access to past data and aggregating it to an area of concern and potentially even analysing this data from a national perspective if desired. Clear structuring of existing data and providing easier access to wider group of specialists could even allow for the future integration of other datasets to give more rigorous insights.

However, it has to be remembered that the AGS data provides a complex granular dataset that needs to be interpreted in the context of other datasets and cannot be used as a standalone dataset to assess risk posed by land contamination. The AGS data format is not an industry data storage standard for geotechnical and geo-environmental data, as it is only meant for transfer. Hence, the use of it in its current state limits the potential for larger data-driven queries and analytics.

There are a few things that should to be considered when analysing large AGS datasets:

- The project ID and location ID need to be merged to generate a unique exploratory hole location ID.
- Due to the nature of the data collection which is done manually by specialists in the field and/or in laboratories, there are errors and inconsistencies.
- Data collection is not automated or carried out by data specialists. Hence, some of the fields that would ideally contain only numeric values may contain text.
- There is great variation in names of chemical compounds (benzo(a)pyrene, Benzo(a)pyrene, benzo[a]pyrene, etc.). The AGS structure assigns unique codes to chemical compounds and groups of chemical compounds. However, the use of these codes is not consistent across the industry.
- Guidance on producing AGS data is generally unclear which can lead to varied qualities of datasets across the industry and slow uptake of best practice.

Potential Framework

5

5.1 Introduction

The third challenge statement the team tackled was to ‘create a framework for encouraging data sharing across built environment professionals’. This was created by utilising the knowledge gained and pain points identified during the previous stages. This allowed the team to identify four key areas for a framework – Education & Collaboration, AGS Data Format, Open Data and Anonymised Datasets. This could help to open-up data on brownfield sites and prepare the industry to make better use of this data to improve safety and deliver better value for our clients.

5.2 Education and collaboration

In recent years there has been a renewed focus within the built environment industry on creating multi-skilled engineers rather than siloed discipline experts as projects become increasingly complex with collaboration and coordination becoming the principal challenges. This discovery project recognises that even solely focusing on the ground, the same is true. We find that to utilise data properly, an engineer or specialist is required to have domain expertise, data science and computer science skills. These three skill sets are quite disparate in the industry and finding the person with the right level of each skill is difficult. Considering typical geotechnical and geo-environmental teams, only a handful of people possess the right skills and are aware of the benefits which could be gained by exploiting data in the way we propose.

We recognise the need for a renewed effort throughout education and industrial institutions to promote data literacy and the diversification of skill sets needed to work with data. This would not only enhance the data generated by a specific part of the built environment, but would also facilitate innovation throughout, and potentially beyond the built environment into other sectors. This builds upon the ODI manifesto for sharing engineering data, which argues that ‘data literacy and skills need to increase across the engineering professions and organisations’¹. Furthermore, it reiterates their point that data literacy needs to be embedded throughout

¹ <https://theodi.org/article/engineering-data-for-the-public-good-a-manifesto/#1565084667757-5e97b880-e130>

the industry, including Universities and Research Intuitions, Professional Bodies and internally within private companies themselves. By making data literacy more prevalent within this industry, the value of data and understanding it becomes more widely known and these benefits can then be passed onto clients.

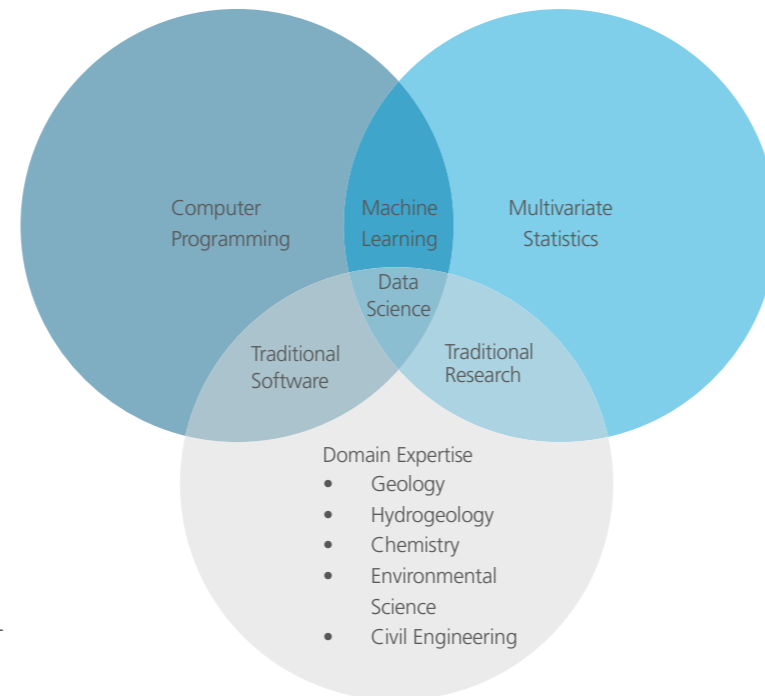


Figure 11 - Geo-environmental Skillset Venn Diagram

5.3 AGS data

In order to facilitate better queries and analytics on bigger datasets, there also needs to be some discussion on the AGS data structure with various stakeholders within the land contamination business (including the AGS Committee, laboratories, contractors, consultancies, regulatory bodies, the BGS and industry bodies). This should be undertaken to highlight the benefits of a unified data structure and agree on robust standards for AGS data. This is to benefit the geotechnical and geo-environmental industry and its stakeholders, enabling more efficient data management and the production of more reliable large datasets.

The sharing of AGS data should also be encouraged by consultants working on land contamination projects. This should be considered at the contract stage and appropriate terms and conditions included and agreed with all parties involved to ensure mutual benefit from opening-up data. This could be done within the BGS *Dig to Share* initiative and appropriate licensing should be associated with the data. The BGS Dig to Share initiative aims to develop a fully digital workflow for sharing data on ground information. For more information see Appendix D.

National AGS standardised datasets would enable a comparison of concentrations of contaminants against GAC and could also unlock additional value, by potentially facilitating the following:

- The derivation of geotechnical and other parameters based on large datasets (this is currently undertaken using very limited or academic datasets).
- Allowing observations on the natural attenuation of contaminants to be made and facilitating the assessment of the efficiency of various remediation techniques.
- Utilising machine learning for land contamination risk assessments.
- Efficient integration with scientific and engineering tools and applications for modelling physical processes (BGS Groundhog).
- Robust integration with interpreted and other project-related data.

AGS Data Standardisation

Atkins has previously identified that the AGS data structure is not unified between different contractors and laboratories involved in producing key aspects of AGS data. Members of the project team therefore produced a technical note to support update of GD07 guidance on “Handling Chemical Data” published in 2011¹ and a specification that can be handed to contractors and laboratories prior to commencing work.

At the time of writing of this report, AGS is not an industrial data format standard, it is considered to be a data transfer format. It’s up to every company to devise their data management plan and manage data according to their own procedures.

We realise that the lack of standardisation potentially limits the potential for efficient data aggregation and data-driven approaches and analytics within the geotechnical and geo-environmental industry. We will therefore engage in a discussion with various stakeholders on this topic.

¹ <http://www.agsdataformat.com/datatransferv4/rules.php>

5.4 Open data

Landowners can be reluctant to share data on concentrations of chemicals within the ground due to cultural issues, liability associated with land contamination and potential impact on land prices. Datasets, such as historical land-uses, that help experts intuitively assess sites for potential contamination, are closed and need to be purchased. Declaring the land quality is part of the planning process, but not all of this data is reported in formats that are computer readable and it is only published temporarily.

London’s 35 local planning authorities/city hall and Atkins are involved in a “London Development Database Automation Project”. This project aims to collect planning applications in machine readable fields to allow for automation and if made available via an open license this data could lead to many new forms of innovations, as was the case with TfL’s open data approach. If a similar initiative could be started to digitise historic data to make it machine-readable too, such as boreholes records or historic maps, even more insights could be extracted for potentially unlocking brownfield sites. Such initiatives are hugely beneficial and would ideally be expanded to include nationwide data, and not just London.

While this level of data is not currently available for land contamination, other areas of the built environment do make their data more openly available. For example, data from within Energy Performance Certificates in England and Wales is made available for individual homes, revealing the

performance and general makeup of over 10 million homes. This is provided in a such a format that makes simple queries of the data possible, including an open access API.

In California, the Water Board requires data on groundwater quality to be shared openly within a GeoTracker database. This contains records for sites that require clean-up. These examples of open data allows people to examine surrounding areas in more detail and provides more confidence in early stage analysis.

Making data more openly available to a wider audience provides a greater opportunity to initially examine the best sites for development, such as brownfield sites. The least relevant sites can be filtered out, allowing a very large number of potential sites to be reduced to a more manageable number that can then be individually interrogated. Allowing a very large number of potential sites to be reduced to a more manageable number that can then be individually interrogated. Furthermore, access to more open data allows policymakers to make more informed decisions, by adding more locational intelligence to a specific site. This process of making data more open would require discussions with key stakeholders to enable this free data sharing on land quality, starting the conversations with how they can open-up data and contribute towards an open data ‘future’ and the benefits this would provide to the industry. These examples of where data has previously been openly shared should reassure key stakeholders.

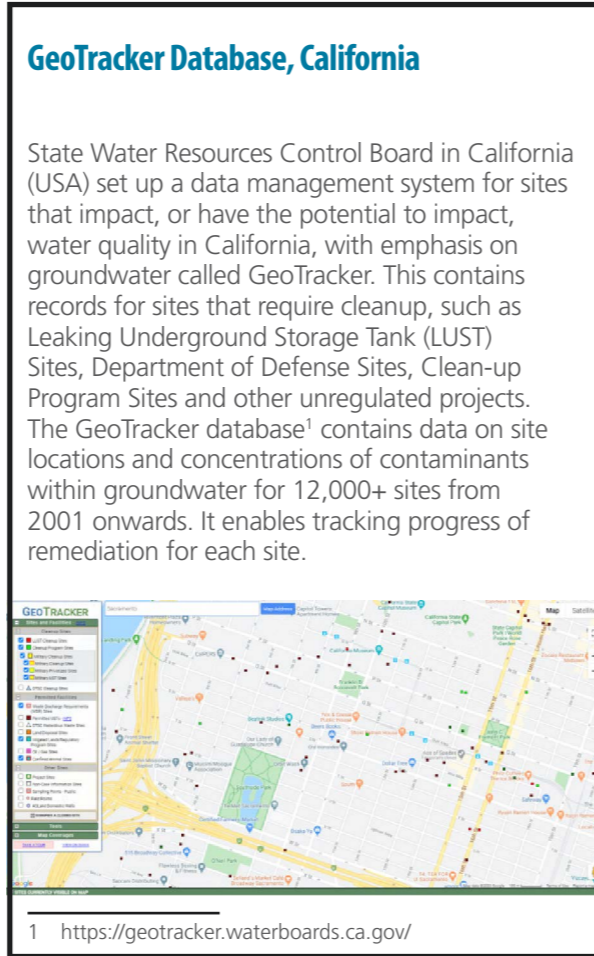


Figure 12 - GeoTracker Database

5.5 Aggregated datasets

To provide confidence to key stakeholders that individual levels of contamination cannot be associated to a specific site we propose that any data that is made publicly available is acceptably anonymised in a similar way to socio-economic data that is typically aggregated to a small area. For example, the Office of National Statistics (ONS) typically aggregates data, such as household income, to small areas with a similar population size like a parliamentary constituency or an output area, see Figure 13. This allows for the mapping of data to geographic areas, without revealing individual situations, so that reliable levels of analysis can still be performed. With this data being geolocated, this then provides an opportunity for it to be combined and analysed with other datasets, to potentially reveal new insights.

Hexgrids (or other) could be used to aggregate data to areas with estimated risk posed by land contamination. This could be generated using a combination of datasets to enable an initial first pass for developers to predict the impact of land contamination on their development and better understand the costs associated with developing the site. This would be useful for initial high-level risk assessment for the development of small brownfield sites and also for large schemes such as rail and/or road schemes. A more detailed land contamination risk assessments could then be undertaken for the proposed design. Potentially, there could also be scenarios where private practices develop an agreement to share the restricted data between

themselves, with the agreement of the client. This would allow the data to remain anonymised, while allowing for greater value to be extracted with restrictive access. The main point is to provide greater access to data that can help with the decision-making process to unlock brownfield sites. Accessing this data should not be a siloed action by individual private companies, it should be coordinated and managed throughout by a governing body, so that it can be strategically planned for the whole industry. This provides organisations (e.g. the Geospatial Commission) the opportunity to govern how data is better accessed and support companies seeking innovative ways to access new data for improving decision making throughout the industry.

Research into this anonymised data should be undertaken and include the following.

- Establishing datasets that could be used to generate the anonymised areas (e.g. historical land use, receptor information, concentrations of contaminants) and the appropriate size of the area.
- A method to validate data before entry.
- The required distribution or density of point data.
- Any statistical limitations of anonymised data and certainty.
- The impact of remediation on the outputs.
- Engaging with key stakeholders (e.g. data owners including Ordnance Survey, Landmark and Groundsure, developers, regulatory bodies, landowners, ODI).



Figure 13 - Output areas diagram for Enfield Local Authority showing ONS’s Output Areas. Each area, represented by a different line weight, has a similar estimated population making them comparable

5.6 Framework summary

This framework sets out four main goals which the project team have established to encourage data sharing across those development companies looking to use brownfield land. These four goals targeted the main pain points that the team identified whilst working on this discovery project. It is hoped that by addressing these, the industry can work together to adopt better practices, potentially opening-up valuable data on brownfield sites and easing the process of redevelopment.

In addition to the four goals, the team have also created a data flow schematic for working with geospatial data in the future. Whilst this is relatively immature, we hope that this will allow others to take this framework and schematic and employ similar techniques themselves. This schematic can be seen in Figure 14.

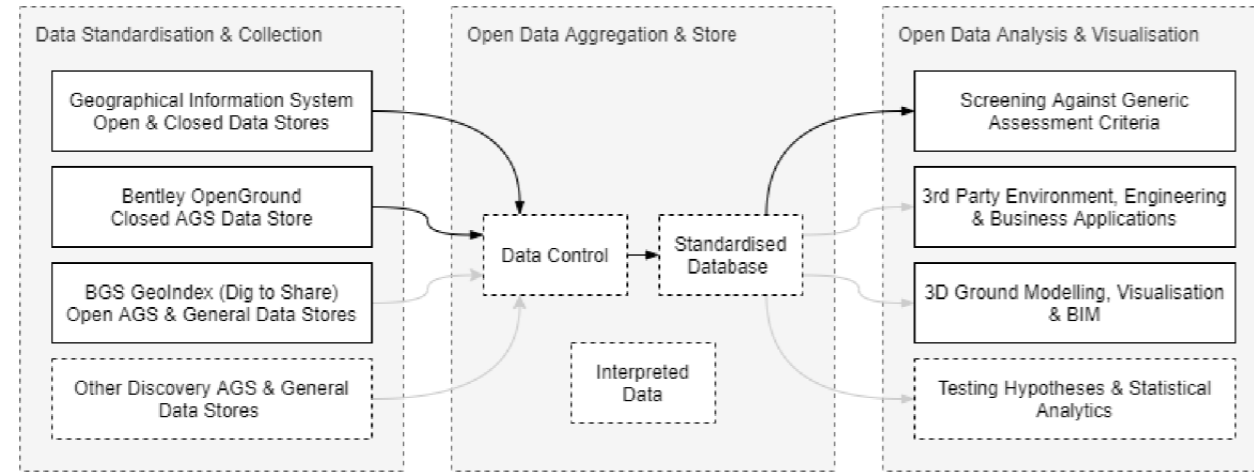


Figure 14 - Planned data flow schematic. Note the solid boxes are the 'mature' elements and the dashed are those 'underdevelopment' with black arrows being connections 'made' while grey arrows are still 'conceptual'

Summary & Next Steps

6

6.1 Summary

In the recent UK Geospatial Strategy, Lord True emphasised that “looking ahead, we must ensure we have a clear strategy to harness the huge potential of location data in helping underpin and drive the UK’s future economic recovery and growth. Location data will be at the fore, empowering and supporting individuals and society.”¹ It is hoped that this will support the government’s aim to “build a million new homes to provide affordable housing to the people who need it, where they need it”² which would clearly have a huge impact on the current social housing waiting list.

¹ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/893455/Geospatial_Strategy_1920x1080.pdf
² https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/893455/Geospatial_Strategy_1920x1080.pdf

Geospatial Strategy 2020-2025

In June 2020, the Geospatial Commission released the Geospatial Strategy 2020-2025. This document outlines four primary missions for the Geospatial Commission as shown below. These missions align very strongly with the outcomes of this discovery project.

- Mission 1: Promote and safeguard the use of location data. We need to provide an evidenced view of the market value of location data, set clear guidelines on data access, privacy, ethics and security, and promote better use of location data.
- Mission 2: Improve access to better location data. We will streamline, test and scale the development of new and existing location data ensuring it is findable, accessible, interoperable, reusable and of high quality.
- Mission 3: Enhance capabilities, skills and awareness: To achieve our vision we must develop more people with the right skills and tools to work with location data, across organisations and sectors, to meet the UK’s future needs and support global development.
- Mission 4: Enable innovation: We will maximise the commercial opportunities for innovation and promote market-wide adoption of high value emerging location technologies.

Data is most valuable when it is used to support decision-making and drive action. In order to achieve this, a sufficient quantity and quality of data needs to be secured. The purpose of this project was to investigate this collection of data and discover to what degree the availability of data would facilitate decision making when developing brownfield sites for housing.

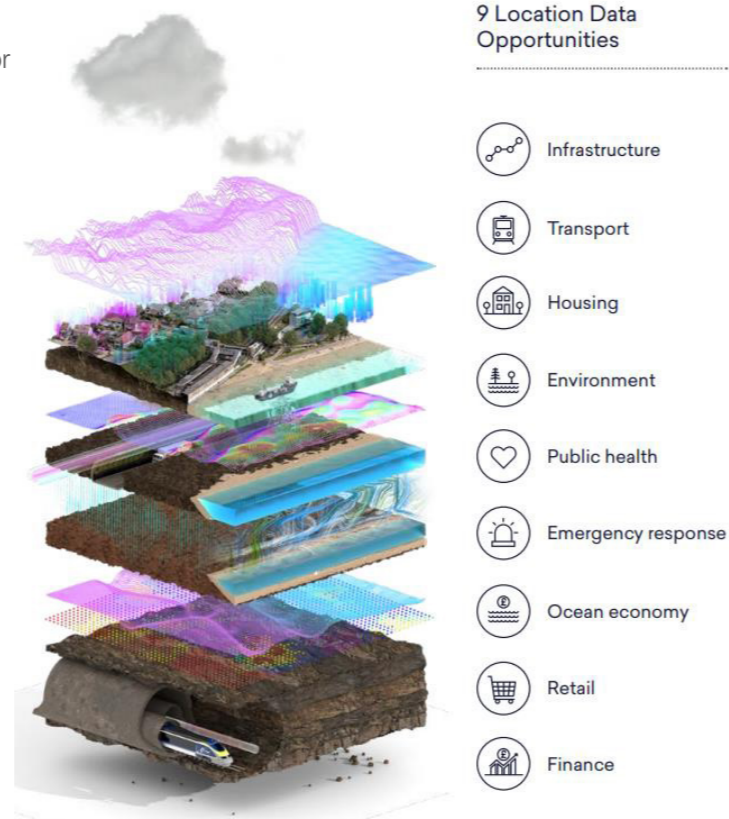


Figure 15 - Location data opportunities (Geospatial Strategy 2020-2025)

6.1.1 Challenge statements

1. Accurately catalogue data that could support understanding of whether brownfield land has been impacted by contamination.

At the scoping stage of this project, the team created a data catalogue which listed a wide range of sources of information associating the impact of contamination on brownfield land. This inventory provides the basis from which a much larger documentation of data could grow and for a system of scoring the data to evolve over time. This has been made possible by creating an open access repository via GitHub, where people can expand and refine the inventory over time.

Not only does this provide a platform for a much more detailed inventory to grow organically, but it also provides the starting point for anyone interested in relevant land contamination data, to see if this could provide insight into a specific site.

To address this challenge, the project team also investigated the AGS data format and discussed its suitability for use. The AGS data format is considered to be a data transfer format, not a data standard. This format is particularly useful for storage of both geotechnical and geo-environmental data as it allows for a wide range of information to be catalogued together. The data within this format offers a granular picture of a specific site.

2. Use case studies to demonstrate how open and closed data of previously developed sites could reduce ground condition uncertainty.

For the case study stage of the research, the project undertook two case studies, a site-specific and a national case study.

The site-specific case study compared the openly available data with the closed data acquired from Groundsure. It was found that the open data didn’t contain enough information to correctly identify the risk of contamination, whereas the private data contained a more complete picture. This requirement for data purchase increases the cost of the risk assessment. This can prove problematic, particularly for small developers.

To address this imbalance the national case study investigated how the AGS data could be potentially opened-up to provide a general understanding of risk posed by land contamination. Whilst inconsistencies exist in the AGS data format and the data needs to be interpreted by a geo-environmental specialist, the study did show that, for a relatively controlled workflow, an indication of the impact of land contamination for a site could be estimated.

3. Create a framework for encouraging data sharing across the brownfield development business.

To address the lessons learned and additional challenges the project team identified during the first two stages of the project, the team created a framework for encouraging and better utilising data sharing related to land contamination. This framework consists of four goals, which are discussed at length in the framework section of this report. The first of these relates to education and institutional support, of which there needs to be more breadth. The other three goals relate to the data itself and include: standardisation of AGS data format; opening-up of data, particularly relating to the ground; and anonymising of data to protect individual sites.

It is hoped that with the implementation of these four practices, data sharing in the industry would be greatly improved and this would undeniably aid the overall ambition of unlocking brownfield sites for housing and other uses. This data would support decision-making and drive action in the industry when it is most needed.

6.2 Key next steps

This discovery project and the conclusions drawn from it have identified a number of valuable next steps for further work and research into this topic that the project team are keen to address:

- Engagement with the Geospatial Commission, working with them to align our research with their four missions outlined in the Geospatial Strategy 2020-2025. These goals will likely be aligned to the development of housing specifically.
- Working with the industry bodies to improve AGS data formatting and opening-up the AGS framework to make it more accessible to a wider audience.
- Initiating conversations with our clients about how they can open-up data on their sites and contribute towards an open data 'future'.
- Further research into the other datapoints available in the AGS data.

List of successes

- The creation of the first open access land contamination data inventory.
- Atkins' first public GitHub repository.
- Code developed to access multi-project data via OpenGround Cloud Web API to allow for large scale analysis of existing data.
- Initial discussions started with key-stakeholder in this sector including AGS, Geospatial Commission and Bentley.
- Multi-discipline research collaboration within Atkins across three disciplines that would not ordinarily work together.
- Many, many great discussions about data...

6.3 Final thoughts

As data becomes more and more prevalent in the built environment and the ambition of a national digital twin gets closer, there is a requirement to start joining different datasets together to add extra locational intelligence to our decision-making processes. However, while these processes may be starting above the ground and build upon numerous existing open data sources, below-ground data, including land contamination, is much more difficult to navigate. Nonetheless, this research project has discovered that there are many great sources of data to create a clearer picture of what's happening below ground. However these are either locked in proprietary sources or require domain specialists to interpret them. Even if this data were to become more accessible, data literacy in the industry is very low and there are fears by many within the industry about sharing data publicly, concerned this may reveal sensitive information..

As a result, this research set out a framework that aims to start addressing these issues and is a framework that aligns itself with the ODI's manifesto for opening-up data in the engineering sector. Alongside this, the research has produced many valuable outcomes, which are often hard to capture in a standard report format. This includes the impetus needed to create an open access data inventory and repository (which is still very rare for large consultants), along with the code necessary to access a large resource of existing data. It has encouraged the team to initiate discussions with key stakeholders, such as the AGS Committee and Bentley. Above all,

it has provided us with an opportunity to challenge the norm and start disruptive conversations within the industry about how we access data and where the true value of this lies.

This is important, because even if there are many hurdles to overcome, below-ground data and specifically data relating to contamination, needs to be a part of this ever-evolving data ecosystem for the built environment and should not be siloed, as is often the case. By making this data as accessible as above ground data, many of these obstacles could be more easily addressed, providing the incentive for better data literacy and trust in sharing data. These types of initiatives, working alongside government strategies, such as the Geospatial Commission's 5-year strategy, would not only aid the industry, but would also provide better locational intelligence to clients by improving decision making. This would ultimately aid the development of underutilised sites, such as brownfield land, for housing and other uses making them more desirable to smaller developers.

Appendices

A What is a Brownfield Site

In the UK, a brownfield site is generally understood as “previously developed land”¹ that has the potential for being redeveloped. It is often (but not always) land that has been used for industrial and commercial purposes and is now derelict². Recently, brownfield sites have become increasingly popular, especially due to the ‘lack of available green spaces for development and in places where demand for residential and commercial property is high’³. For example, around a third of all brownfield sites in England ‘are in the high-growth areas of Greater London, the South East and East’.

The UK is committed to developing brownfield sites as a priority. This led to the introduction of “The Town and Country Planning (Brownfield Land Register) Regulations 2017”⁴ that require local authorities to prepare and maintain registers of brownfield land that is suitable for residential development⁵. These Brownfield registers are meant to provide up-to-date, publicly available information on brownfield land that is suitable for housing via local authority online outlets. This aims to improve the quality

¹ This is an article that describes brownfield sites, their definition, assessment and remediation, and their relevance to sustainable development <http://www.sustainablebuild.co.uk/brownfieldsites.html>

² <http://www.sustainablebuild.co.uk/brownfieldsites.html>

³ <http://www.sustainablebuild.co.uk/brownfieldsites.html>

⁴ <http://www.legislation.gov.uk/uksi/2017/403/contents/made>

⁵ The Town and Country Planning (Brownfield Land Register) Regulations 2017 <http://www.legislation.gov.uk/uksi/2017/403/contents/made>

and consistency of brownfield land data held by local planning authorities and provide certainty for developers and communities, encouraging investment in local areas and specifically underutilised brownfield land⁶.

The first part of the brownfield registers is a comprehensive list of all brownfield sites in a local authority area that are suitable for housing, irrespective of their planning status. These, registers are also meant to be a ‘vehicle for granting permission in principle for suitable sites where authorities have followed the relevant procedures’⁷. If the authority considers that permission in principle should be granted for a site, the local authority is required to enter that site into the second part of the register⁸. Part 2 is therefore a subset of Part 1 and will include only those sites for which have permission in principle granted. However, a site may not be included on Part 2 of the register where development of the site would:

- fall within Schedule 1 of the Environmental Impact Assessment Regulations;
- has been screened as Environmental Impact Assessment development; or

⁶ <http://www.sustainablebuild.co.uk/brownfieldsites.html>

⁷ <http://www.legislation.gov.uk/uksi/2017/403/contents/made>

⁸ This is the national data standard that all local planning authorities in England are encouraged to follow in preparing and publishing their brownfield land registers. <https://www.gov.uk/government/publications/brownfield-land-registers-data-standard>

- development would be prohibited under habitats protection legislation, i.e. those sites may not be granted permission in principle through being placed on the register.

Permission in principle is meant to settle the fundamental principles of development (use, location, amount of development) for the brownfield site giving developers/applicants more certainty. However, a developer cannot proceed with development until they have also obtained technical details consent⁹. This is a new type of planning permission design to address specifically the need for more homes in brownfield sites.

The Campaign to Protect Rural England (CPRE) in their “State of Brownfield 2019”¹⁰ report clearly highlight that the ‘redevelopment of suitable brownfield sites can clearly go a long way in delivering the homes the country needs in places people want to live’¹¹. However, they also discuss how ‘more needs to be done to ensure that brownfield registers deliver on their potential to proactively encourage delivering new homes on suitable brownfield land’¹². Hence,

⁹ <http://www.legislation.gov.uk/uksi/2017/403/contents/made>

¹⁰ <https://www.cpre.org.uk/resources/state-of-brownfield-2019/>

¹¹ CPRE’s annual State of Brownfield report shows that there is enough suitable brownfield land available in England for more than 1 million homes across over 18,000 sites and over 26,000 hectares. <https://www.cpre.org.uk/resources/state-of-brownfield-2019/>

¹² <https://www.cpre.org.uk/resources/state-of-brownfield-2019/>

they recommend that Local Planning Authorities should ‘follow the data standards more rigorously and publish an aggregated national brownfield register to support communities and developers in identifying which sites are on the brownfield register and help identify any gaps’¹³. This would prevent the need for organisation like the National Housing Federation to combine over 340 individual registers (often in different formats, column orders, using different coordinate systems and inconsistent data inputting) and allow the national picture to be openly available to all.

Furthermore, they recommend that the Brownfield Registers Regulations should continue to evolve, so that ‘registers act as a true pipeline, identify all brownfield sites and record their suitability for uses other than housing as well as recording their ongoing development status and the contribution they have made towards meeting development needs after completion.’

On a similar note HTA Design LPP (HTA) also

¹³ <https://www.cpre.org.uk/resources/state-of-brownfield-2019/>

completed a report called ‘Unlocking Potential: Best Practise for Brownfield Land Register’¹⁴ documenting the accuracy of Brownfield Land Registers. From this they frankly stated that ‘collation of the register has not led to the production of any new information, or the adoption of any new procedures/tools, to identify new brownfield sites for development’¹⁵. Also, they highlighted the fact that ‘the search for, and identification of, developable brownfield sites is currently largely done by officers within the planning policy department of local authorities’¹⁶ and they advise that wider skills are required to make this process efficient. HTA research also highlights that at present ‘there is uncertainty around the purpose of the Brownfield Land Register’, questioning ‘whether it is to provide accurate data on the amount of brownfield land available for redevelopment, or whether it is a planning mechanism which can be used to provide certainty to developers that brownfield sites are suitable for housing.’

¹⁴ <https://www.cpre.org.uk/resources/unlocking-potential-best-practice-for-brownfield-land-registers/>

¹⁵ On behalf of CPRE, HTA Design LLP has looked into the process local planning authorities undertake to identify brownfield sites for development. The report sheds light on some of the biggest challenges facing local authorities when compiling Brownfield Land Registers. <https://www.hta.co.uk/news-description/second-news>

¹⁶ <https://www.hta.co.uk/news-description/second-news>

B Land Contamination Risk Management

Within part 2 of the brownfield land register, where permission in principle is granted, there is a section that requires Hazardous Substances within the land to be acknowledged. It requests that the local planning authority conducts an environmental impact assessment on the proposed development as per regulation 26(3) of the Planning (Hazardous Substances) Regulations 2015.

All brownfield sites need to be assessed by “a suitably qualified competent professional” before they can be redeveloped as per “Land Contamination: Risk Management” (LCRM)¹⁷ guidance and Contaminated Land Report 11¹⁸ (to be withdrawn after consultations for LCRM are finished). This is to assess the risk to environment and people. The guidance covers:

- assessing the risks
- making appropriate decisions
- taking action where necessary

The guide is used in a range of regulatory and management contexts such as voluntary remediation, planning, assessing liabilities or under the Part 2A¹⁹ contaminated land regime.

¹⁷ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

¹⁸ <https://webarchive.nationalarchives.gov.uk/20140328160926/http://cdn.environment-agency.gov.uk/scho0804bibr-e-e.pdf>

¹⁹ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/223705/pb13735cont-land-guidance.pdf

Risk management comprises of three stages – risk assessment, options appraisal and remediation. Data is collected and assessed during all three stages and all of them comprise of three tiers. Risk assessment is the most data-heavy process.

Upon the completion of each tier of risk assessment, a conceptual model is developed/re-defined. Conceptual model contains information on potential sources of contamination, pathways and receptors. Where all three exist, a potential pollutant linkage is identified, and the risk can be deemed unacceptable. The next step in the risk assessment is undertaken only when a risk in the previous tier has been unacceptable.

- Tier 1: Preliminary risk assessment (PRA) – information/data is collected during this stage from a site walkover and via a desk study, risks are assessed qualitatively;
- Tier 2: Generic quantitative risk assessment (GQRA) – this stage builds on data assessed during PRA, data is typically collected on site during site investigation. Each potential pollutant linkage is then reassessed by comparing the contaminant concentrations against appropriate generic assessment criterium (GAC);

- Tier 3: Detailed quantitative risk assessment (DQRA) – further detailed site-specific information is collected and site-specific assessment criteria (SSAC) are developed, some form of (numerical) modelling is typically undertaken.

Should risks be considered unacceptable, remediation needs to be undertaken. Options appraisal is undertaken first to identify feasible remediation options followed by selection of the most appropriate solution. More risk management for land contamination can be found in LCRM²⁰.

Remediation can be expensive and complex, and this needs to be considered before purchasing brownfield land. Not all sites are deemed feasible to be remediated, particularly if the costs exceed the value of the land after development²¹. This is often a major

²⁰ <https://www.gov.uk/guidance/land-contamination-how-to-manage-the-risks>

²¹ <http://www.sustainablebuild.co.uk/brownfieldsites.html>

blocker for small to medium sized sites, where the financial risk can often outweigh the potential return.

It should also be noted that contractors undertaking ground investigation have to comply with Construction (Design and Management) Regulations 2015 and Guidance Notes for the Safe Drilling of Landfills & Contaminated Land (British Drilling Association, 1992). Prior to any ground investigation, a desk study for the site is to be undertaken to identify any potential source of contamination. Sites are then classified as Green, Yellow or Red depending on the level of contamination.

C OpenGround Cloud

Bentley OpenGround Cloud is a proprietary platform of integrated applications for easily and robustly entering, collecting, managing, visualizing, analysing and reporting ground (AGS) data. It is designed primarily for project-by-project data management, due to traditional project and client requirements, providing teams across the project supply chain controlled access to a central data source via connected applications to increase project collaboration, efficiency and assurance. It is a solution used widely as best practice within the industry by multiple consultants and contractors.

D Dig to Share

The Dig To Share project is a collaboration between Atkins, Morgan Sindall and the British Geological Survey (BGS) and focuses on sharing ground investigation data across the industry. The project secured £50,000 funding from the Infrastructure Industry Innovation Platform (i3P) in April 2018 when Sophie Payne (Atkins) and Holger Kessler (BGS) pitched at a Dragons Dens style event 'Spark'.

The project aims to drive a culture change and get more people working together to unlock our ground data and create a fully digital workflow accessible to the whole industry. This would use existing BGS systems along with new connections through data management software, such as with OpenGround Cloud. During the early stages of the project several collaborative workshops were set up to discuss the opportunities and challenges. Fluxx, as part of our Digital Incubator team, led a behaviours study interviewing contractors, consultants and clients to identify blockers and develop solutions.

In 2019 the project really gained momentum, with members of the project team presenting and exhibiting at several conferences across the UK to promote the project. The feedback from these events was very positive, with lots of encouraging feedback regarding the project and the need for information to be shared, open source and accessible.

The project has now developed a community of 50 Super Users from across the industry, including clients, consultants and contractors with the common goal to collaborate and raise awareness of the benefits of using and adding data to the BGS database. The community has allowed for the sharing of best practice, tips, ideas and queries; distributing these amongst colleagues within their organisations. The Dig To Share team have produced a Super User Welcome pack to support the Super Users with their roles and have also presented several online webinars which have showcased the BGS database and the data ingestion tools, including the built-in tools within OpenGround Cloud.

The adopted target was that 10,000+ boreholes would be released in AGS format in the first year and that a community would be developed around open data and the creation of a simple sustainable workflow.

In July 2019, the project reached the 10,000 AGS borehole release aim and has created an active self-perpetuating network of Super Users for data sharing. The project so far has helped to release the Crossrail data and gain access to Network Rail, the Environment Agency and Highways England data, with data sharing agreements in place with all of these client organisations. In 2019 the project was "Highly Commended" at the Keynetix Geotechnical

Data Management Awards and in 2020 the project is one of the finalists for "Project Team of the Year" for the Dig To Share project. Later this year the project team will release information on a new Citizen Science Initiative to digitise some of the 1.4 million historical boreholes within the BGS database.

The i3P funding ended in June 2019 and since then the project has been continuing with the voluntary time of members from across the Dig To Share team along with our Super Users. The strong relationships built between members of the team has shown to be very effective in seeing the project surpass its targets and look to continue in the next project phase and on future collaborations.

For more information on the project, information on the workflows and how to join as a Super User visit the Dig To Share website: <https://i3p.org.uk/2018/08/22/dig-to-share/>

ATKINS

Member of the SNC-Lavalin Group

Caroline Paradise

Atkins Ltd
Nova North
11 Bressenden Place
London
SW1E 5BY
UK

caroline.paradise@atkinsglobal.com