Improving data and information access in regulated sectors

Exploring data and information access and innovation in the regulated banking, accounting and legal sectors

Open Data Institute

Contents

Foreword 3

Executive summary 6

Background 9
- Data access and innovation in regulated sectors 10
- Data access and innovation in the banking sector 12
- Data access and innovation in the accounting sector 12
- Data access and innovation in the legal sector 13

User research findings 14
- People insights from market research 14
- Process insights from market research 15
- Technology insights from market research 18

Recommendations 20
- For data intermediaries 20
- For data providers 21
- For government, regulators and industry bodies 22

Appendix 1: Methodology 23

Appendix 2: Additional Resources 26

About

This report was researched by the Open Data Institute (ODI) and commissioned by RegulAItion. The project is funded by UK Research and Innovation through the Next generation services challenge. It was published in July 2021. The lead authors are Josh D'Addario and Arindra Das, with input from Olivier Thereaux, ODI.

To share feedback by email or to get in touch, contact the ODI, at info@theodi.org, or RegulAItion, at contactus@regulaition.com

Find information on the data access programme here. The aim of this competition was to speed up the responsible adoption of artificial intelligence (AI) and data technologies and solutions in the accountancy, insurance and legal sectors by enabling better access to data. UK Research and Innovation will invest up to £3.5 million to this effect.

How can it be improved? We welcome suggestions from the community in the comments.

Foreword

by Sally Sfeir-Tait, RegulAItion CEO

The fourth industrial revolution is upon us[1]. Governments understand this deeply[2] and have devised strategies to support the transformation of their economies[3]. If they do not, they risk seeing their existing industries replaced by technology providers, likely located in other countries[4].

All sectors are experiencing a fundamental transformation. Technology products and services are testing the traditional boundaries of technology and financial and professional services. This is mostly apparent in the context of Big Tech with companies like Alibaba, Tencent, Alipay[5], Google, Facebook and Amazon[6]. This has prompted the coining of a new term, TechFin[7], and the development of numerous studies on the impact[8] of technology companies delivering financial services[9] and the fundamental questions we must consider[10].

Many in the traditional industries of financial and professional services still believe that transformation is happening in the 'other' sectors, not in theirs. While some believe that transformation has already happened and that they are too late.

A few, who have taken the time to understand existing disruptive technologies know that they are masters of their own destiny. By leading the transformation, they ensure that they secure new revenue opportunities for their organisations and seize operational efficiencies today[11]. We, at RegulAItion, are lucky that these individuals and the organisations they represent have chosen us as partners.

They have chosen the AIR Platform to deliver access to information as they build a new knowledge management infrastructure to support all of their digital transformation initiatives.

The challenge we are all facing today is one of language and communication. We hear talk of 'data access', 'data sharing', 'data availability', 'information sharing', 'knowledge management' and many more expressions and terms that all relate to the same subject: managing the increasing volume of knowledge we are all individually and collectively creating. How do we work together to gather and coordinate the knowledge we have and use it to serve our collective and common interests?

---

[1] WE Forum (2016), 'The Fourth Industrial Revolution: what it means, how to respond'
[2] GOV.UK (2017), 'Growing the artificial intelligence industry in the UK'
[3] UK Research and Innovation (2021), 'Next Generation Services Grand Challenge'
[4] HM Treasury (2017), Regulatory Innovation Plan
[5] Institute of International Finance (2018), 'Bigtech in China'
[6] Amazon providing credit to merchants on its marketplace, Google's tie up with CitiBank to provide account services and Facebook's pursuit of launching its own cryptocurrency.
[7] My Tech Mag (2020), 'FinTech vs. TechFin: Where is the Future Headed?'
[8] Financial Stability Board (2019) 'BigTech in finance: market developments and potential financial stability implications'
[9] Institute of International Finance (2018), 'A new kind of conglomerate: Bigtech in China'
[10] European Banking Institute Working Paper Series 2017 - No. 6 (2017), FinTech to TechFin: The Regulatory Challenges of Data-Driven Finance
[11] Mckinsey (2018), 'AI has the potential to deliver additional global economic activity of around $13 trillion by 2030'

The global pandemic has shown that sharing knowledge can be existential to human life and to our economies. The UK government is deeply aware that an infrastructure approach to knowledge management is essential. The National Data Strategy has the potential to confer a unique advantage to the UK, helping to drive innovation, meet ambitious policy goals, secure trade deals, address systemic risk and create a new competitive advantage on the world stage.

Multiple technologies and solutions are coming to the rescue. How do we decide which technologies to adopt and why? That's why this report is so important. It shares knowledge about the state of the industry and recommends actions in a clear framework that can be followed by businesses, regulators, governments and industry bodies to improve how we access information and share knowledge in regulated sectors.

This is an important and timely message. Breakthroughs for humankind have, after all, always come about through sharing and collaboration. Sharing of knowledge dates back to the earliest cave paintings. Societies evolved by communities working to create a common purpose. As we now face new frontiers, it's time to reimagine how we share and work together.

Imagine a world where intelligence can be quickly unlocked in order to overcome complex health, economic and environmental challenges. Where competitors can safely work together to achieve common goals.

That is the future each and every one of us at RegulAItion and the ODI can see. It is the one we all want to live in. And this report is crucial in bringing us one step closer to it.

Executive summary

Trustworthy and sustainable data access is the underpinning factor of any healthy data ecosystem. This is true in regulated environments where data is particularly sensitive, privacy and confidentiality are enshrined for the main players, and the reputational and financial risks regarding compliance are high.

The Open Data Institute (ODI) and RegulAItion, funded by Innovate UK's Industrial Strategy Challenge Fund, worked together to explore the landscape of access to data across the regulated sectors of banking, accounting and law. In particular we researched how innovative privacy enhancing technologies (PETs), and developments in machine learning (ML) and artificial intelligence (AI) could improve trustworthy and sustainable data access across these sectors.

Historically, the sectors of banking, accounting and legal have been relatively change resistant when it comes to deploying data access technologies, with much of their technological investment focussed on data privacy and security instead. This is for a number of factors, chiefly that confidentiality and trust are crucial elements of the services provided. In fact, as an industry expert said during our research, "trust is the main product accountants sell".

Despite this deeply embedded culture of privacy and risk aversion, there has been some notable progress. The most substantial and public example has been in the finance sector with the implementation of open banking – a data portability initiative intended to empower banking consumers with greater control over data about their finances, and to reduce the market power of the UK's nine largest banks. The accounting and legal professions have not had such a seismic policy change, but have been investing in ML and AI technologies to augment the skilled work of professionals in these fields. Organisations including The Institute of Chartered Accountants in England and Wales (ICAEW) and LawtechUK are leading in innovative research and development in their respective sectors.

There is now an opportunity for more data innovation across these sectors that would have been unlikely even a few years ago. Findings from our research show that these changes can be seen in three interconnecting ways: capacity building in people, improvements in processes, and advancements in technology.

People in these industries are more positive about data innovation than ever. The increasingly hi-tech demands of clients, and the growing influence of internal champions, have pushed these sectors to reevaluate desired skills. This has led to both an increase in data scientists and other data roles (including ethics and strategy) within established businesses, and increased comfort with bringing in data and analytics companies to perform important functions when there is not the required expertise internally.

Processes for improving data access and sharing towards innovative ends have gone hand-in-hand with technological advancements. Although the need for highly secure

and privacy-preserving activities remains as strong as ever, there have been pushes for more open innovation processes. Examples include hackathons, data challenge series and open forums for sharing best practice and learnings in trusted environments.

Technology has been a key enabler across these changes. The explosion in PETs over the last decade has fuelled trust in more open systems, and created increased demand for data science capabilities to deploy these technologies. From anonymisation techniques, to sandbox environments and synthetic data, to algorithmic privacy techniques like federated learning, PETs are driving data innovation across these sectors.

Although large organisations like multinational finance firms are investing a lot in people, processes and technology in-house, a relatively new role has emerged that is also well placed to provide these solutions to the market, the data intermediary. Intermediaries facilitate the access to data and information across data ecosystems. This is often done by collating disparate datasets in a central location and vetting providers and users of that data. New developments in PETs such as federated learning are also allowing intermediaries to increase information flows in an ecosystem without 'raw' data ever being accessed or shared. This new generation of intermediaries, including RegulAItion, is creating value within their ecosystems without direct data access, improving security, privacy and data governance. When data intermediaries securely facilitate valuable data and information flows in a sector, they help to build a more open, trustworthy and sustainable data ecosystem.

In order for the regulated data ecosystems of the banking, accounting and legal sectors to further realise the benefits of data innovation, through better products and services, improved consumer outcomes and easier legal compliance, data access needs to be improved. Data intermediaries have an important role to play in this, but everyone with a stake in the ecosystem needs to play their part.

Data providers should convene an ecosystem of partners, suppliers and customers to explore ways to increase access to data that will create business benefits, while respecting privacy and minimising potential harm. Running or participating in innovation programmes and accelerators or academic partnerships can be a safer and more trustworthy way of experimenting with data. Partnering with intermediaries that can create value while minimising data exposure can help build secure, trustworthy and sustainable data access models.

Governments and regulatory bodies should fund research and innovation programmes that will help to facilitate data collaboration to deliver value within specific sectors and ensure that any data accessed or shared as part of regulatory changes is made as open as possible within the appropriate considerations toward security, sensitivity and privacy. Professional associations and industry bodies should convene businesses within their networks to help map out current and future data ecosystems that will unlock value for all participants

Greater data collaboration between data providers and data intermediaries within the banking, accounting and legal sectors will create more sustainable and trustworthy

data ecosystems, ultimately driving greater value to businesses, consumers and wider society.

Background

Trustworthy and sustainable data access is the fundamental factor of any healthy data ecosystem. Our research into the value of data sharing in the private sector has demonstrated that more data flowing through a business ecosystem creates benefits for data users, data providers and beyond.[12] Increased data access can improve market reach, identify market insights and optimise supply chains. But without trust, data users will not access data that could otherwise further their goals; and without financial sustainability, that access eventually dries up.

This is especially true in the regulated sectors of banking, accountancy, and law where data is very sensitive, and regulation and compliance are very strict. In these data ecosystems, we see that data flows are restrained both internally and externally due to the needs of legal compliance and fears of breaching regulation. Strict data protection and industry regulation can be a solid platform for ethical innovation in this space that would allow for a more dynamic ecosystem without reducing security and privacy or increasing harm.

One reason for improving access to data is to create value through innovative new technologies. As highlighted in 'Growing The Artificial Intelligence Industry In The UK' by Professor Dame Wendy Hall and Jérôme Pesenti, increased use of AI can bring major social and economic benefits.[13] With AI, computers can analyse and learn from data with higher accuracy and speed than humans alone. AI could improve efficiency and performance in many sectors, including banking, accounting and legal.

Professors Hall and Pesenti identified that in order to acquire these gains, organisations must improve access to data by:

- Facilitating the sharing of data between organisations holding data and organisations looking to use data to develop AI through proven and trusted frameworks and agreements, which would ensure exchanges are secure and mutually beneficial.
- Improving the availability of data for developing AI systems. Government should ensure that public funding for research explicitly ensures publication of underlying data in machine-readable formats with clear rights information, and open wherever possible.
- Supporting text and data mining as a standard and essential tool for research by establishing by default that for published research the right to read is also the right to mine data, where that does not result in products that substitute for the original works

In line with the trends identified in this paper, RegulAItion was launched in 2018 to help unlock intelligence by reimagining the way in which organisations across the

[12] Open Data Institute (2020), 'Report: Sharing data to create value in the private sector', https://theodi.org/article/report-sharing-data-to-create-value-in-the-private-sector/
[13] GOV.UK (2017), 'Growing The Artificial Intelligence Industry In The UK', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

globe access data, without sacrificing privacy or security. RegulAItion's AIR Platform blends advanced federated learning, encryption, blockchain and AI techniques to change the way data is accessed in the regulated sectors. It is developing the technology infrastructure and ecosystem for scalable, automated, repeatable and responsible AI-driven data access, by enabling travelling algorithms to spot patterns in data without datasets needing to leave their secure premises.

To better understand this burgeoning market, the ODI and RegulAItion embarked on a research project across these sectors, exploring the needs, barriers and opportunities related to data access and sharing through AI and ML-based PETs. Our goal was to review existing situations or problems and explore industry knowledge around working with these types of data access and sharing.

This report includes some brief snapshots across the sectors in scope, the user research findings organised by People, Process and Technology, and recommendations for different organisation types across the sectors.

Data access and innovation in regulated sectors

Every business is subject to regulation governing how it collects, uses and shares data.

UK businesses, regardless of industry, need to comply with data protection laws, such as the Data Protection Act, as well as requirements to report health and safety data, pay equality, etc. Specific industries are also subject to regulation and laws that require them to collect and share data with regulators.

For example, businesses across the UK financial services industry need to share financial data with the Financial Conduct Authority (FCA). The Financial Reporting Council (FRC) is responsible for regulating auditors, accountants and actuaries, receiving data from them during audit periods. Legal services in the UK are regulated by a variety of organisations, including the Bar Standards Board in England and Wales, and different organisations for Scotland and Northern Ireland.

Regulatory compliance is necessary for companies to continue operating within their industry; though this is sometimes portrayed as an unwelcome obligation, the need to comply with regulation can be an opportunity to re-engage with customers and to drive innovation.

As we noted in our case study 'The value of sharing data in regulated environments' in 2020, businesses share data both for regulatory purposes and more widely for their business.[14] Beyond being a legal requirement, regulated data sharing often creates benefits for the organisations that share the data, as well as others in their industry and consumers. Often the most value is created when organisations take anticipatory actions or are 'first movers' in a newly regulated space, taking advantage of new

[14] Open Data Institute (2020), 'Case study: The value of sharing data in regulated environments', https://theodi.org/article/case-study-the-value-of-sharing-data-in-regulated-environments/

opportunities with limited competition, and demonstrating commitment and trustworthiness to regulators and the wider community.

Data intermediaries have been helping build and grow these data ecosystems for years. They provide a range of services much needed by data providers and not found in-house, predominantly technical, but also networking and advocacy services to ecosystems.[15]

Regulators have an active role to play as well. They are well-situated to help improve how data is used in their sectors and ecosystems by setting norms and standards or supporting data flows.

Three aspects of data infrastructure that are particularly important for supporting innovation, and where regulators could have a role, are:

- access to data – who gets to use it and what for, which might include identifying when data within a sector should be open, or who should be able to access data that is not open
- data governance – what policies and practices there should be to minimise harms and improve trust in data
- data standards – how to ensure that data can be combined in flexible ways, and is interoperable

Organisations like the Better Regulatory Executive (BRE) are focussed on finding the right combination of regulatory interventions to protect and empower citizens while supporting innovation. Its paper '[Regulation for the fourth industrial revolution]' includes a number of commitments to improve access, governance and standards across regulated sectors.[16]

The UK is in a strong global position to deliver on the benefits of more innovation-driven regulation through improved access to data. It is currently ranked first out of 37 countries across aggregate indexes of both primary and secondary legislations in the OECD Regulatory Policy Outlook, fourth in the Global Innovation Index, and joint first among 30 governments that have made concrete commitments to champion open data.

Organisations within the UK's banking, accounting and legal sectors have already begun making significant moves to both improve data access and invest in innovative technologies like AI within, and across, their sectors.

Data access and innovation in the banking sector

A major driver of data access and innovation in the UK banking sector over the last few years has been the implementation of [open banking], and, with expansion into

[15] Web Foundation (2015), "Open Data Intermediaries: Their Crucial Role", https://webfoundation.org/2015/08/open-data-intermediaries-their-crucial-role/
[16] GOV.UK (2019), "Regulation for the Fourth Industrial Revolution", https://www.gov.uk/government/publications/regulation-for-the-fourth-industrial-revolution/regulation-for-the-fourth-industrial-revolution

open finance and the wider smart data ecosystem planned, this trend is expected to grow. Open banking is a data portability initiative designed to increase competition and innovation in the UK's banking market by allowing individual and SME bank customers to securely share their data with third parties so a broad range of businesses can compete to provide bank customers with better financial services.[17]

Access to the increasing volumes of this data available through open banking has resulted in an accelerated use of AI and machine learning to create new propositions, such as in credit applications,[18] chatbots and personal financial managers.[19] Open banking and open finance driven AI may further improve operational efficiency through automated and autonomous decision-making and insight creation.[20]

Another area in the wider finance landscape where data sharing, AI and data intermediaries are proliferating is in alternative data for investing, or 'alt data'.[21] Alt data refers to new, non-traditional types of data, used to inform investment decisions. This may be data about individuals such as footfall data, data about business processes such as credit card spend at retail stores, and data from physical sensors or satellites.[22]

These new data sources generate lots of unstructured data but it is difficult to analyse its value or combine it with other alt data and traditional investment datasets. ML processes are being deployed to decipher the potential value of this data.

Data access and innovation in the accounting sector

Data is also central to the work done by accountants and auditors. Although much of the data that is still used today in these processes is not digital, there have been significant advances in data access and innovation with the knowledge that this can help accountants deliver more value to businesses.

Accountants and auditors have many opportunities to use different sources of data and new analytics tools. Accountants are increasingly drawing on real-time data and non-financial data to drive cost analysis and forecasting, while auditors use data analytics to identify outliers and exceptions, and focus on the areas of greatest risk.[23] This uptick in access to non-traditional data to be combined with traditional sources will continue to improve these analytics as PETs allow for greater access to sensitive financial data.

---

[17] Open Banking Ltd. (2019) "Open Banking, Preparing for lift off" https://www.openbanking.org.uk/wp-content/uploads/open-banking-report-150719.pdf
[18] Lexology (2020) "Open banking, 'big data' and AI: what are the implications for the financial services sector?", https://www.lexology.com/library/detail.aspx?g=688a8145-bb2b-46c6-bcb1-6af52c6e961e
[19] Dulcit.AI (2021), "Conversational AI",https://www.ducit.ai/#conversationalai
[20] Euro Banking Association (2021), "Artificial Intelligence in the era of Open Banking", https://www.abe-eba.eu/thought-leadership-innovation/open-banking-working-group/management-summary-artificial-intelligence-in-the-era-of-open-banking/
[21] ODI (2020), "Alternative data: exploring the ethical implications", https://theodi.org/article/alternative-data-exploring-the-ethical-implications/
[22] ODI (2020), "Building trust in alternative data ecosystems", https://theodi.org/article/building-trust-in-alternative-data-ecosystems/
[23] ICAEW (2019), "Big data and analytics: the impact on the accountancy profession", https://www.icaew.com/technical/technology/data/data-analytics-and-big-data/big-data-and-analytics

ML techniques have existed in accounting and auditing for some time but widespread adoption across the sector is still in early stages. Industry leaders are still developing a full understanding of how AI can solve accounting and business problems. Organisations like the ICAEW have identified areas with fruitful early research and implementation projects in the areas of coding accounting entries, improving fraud detection, forecasting revenues and analysing unstructured data.[24]

Data access and innovation in the legal sector

Lastly, the legal sector also provides a wide variety of opportunities for data access and AI. While legal services have been traditionally characterised as a relatively conservative market, new technologies, or LawTech, are disrupting the market as law firms have adopted new technologies to reduce litigation costs, improve court system efficiency and drive broader innovation. Currently most LawTech solutions focus on routine and administrative work, however investments are being made to achieve more fundamental transformations using ML and AI.[25]

The main market drivers for improving privacy-preserving access to data and AI solutions are the need for greater efficiency, increasing workloads and complexity of work, increasingly 'tech savvy' staff and greater client pressure on costs and speed.[26] This field is still less mature than other adjacent sectors such as financial technologies (fintech) where there is more funding and regulatory alignment. Within the legal sector the B2B market is the most mature, particularly within large law firms, with adoption of AI/ML applications in areas like legal analytics, project management and compliance and contract management.

The UK continues to invest in this space as a growth sector, increasing funding for collaboration between businesses and researchers, AI-based PhDs, and more. Most recently the University of Oxford launched a government-funded research programme to explore the potential and limitations of using AI in support of legal services.[27]

[24] ICAEW (2018), "Artificial intelligence and the future of accountancy", https://www.icaew.com/technical/technology/artificial-intelligence/artificial-intelligence-the-future-of-accountancy
[25] TheCityUK (2019), "Legal Excellence, Internationally Renowned: UK Legal Services 2019", https://www.thecityuk.com/assets/2019/Report-PDFs/294e2be784/Legal-excellence-internationally-renowned-UK-legal-services-2019.pdf
[26] The Law Society (2019), "Lawtech Adoption Research report", https://www.lawsociety.org.uk/en/topics/research/lawtech-adoption-report
[27] Thomson Reuters (2019), "Legaltech Startup Report: A Maturing Market", https://legalsolutions.thomsonreuters.co.uk/en/trends-insights/reports/legaltech-startup-report-2019.html

User research findings

From October 2020 to April 2021 we engaged in user research using a variety of structured and semi-structured interviews and roundtables to collect views on the opportunities and challenges around data access and sharing, PETs and the use of AI and ML in regulated sectors, with a focus on the banking, accounting and legal sectors. User research participants included banks, accounting firms, legal services providers and legal departments, regulators and government, industry bodies and technology policy organisations.

We have organised the findings according to the ITIL framework of People – including skills, attitudes and culture, Process – including organisational approaches, frameworks and policies, and Technology – including technological solutions deployed internally and services purchased externally.

People insights from market research

Low sector maturity in 'open culture' and innovation champions

There are a number of deep rooted barriers that often hold large organisations back when it comes to data access. These include privacy, trust, transparency, commercial and ethical concerns. Perhaps the largest of all is a human resistance to change.

Our research has identified the importance of 'champions' within organisations. Individuals who aren't afraid to challenge the status quo and are keen to push new innovations. These champions tend to have a background in or familiarity with technology as opposed to many of their colleagues. Often, champions from different firms will use networks and competitions to share ideas and best practice, and look for opportunities to develop pilots or case studies to build an internal business case for an innovation project.

There are a variety of organisations providing these services across the different sectors, including Innovate Finance in banking, the Tech Faculty at the ICAEW and DFK International in accounting, and LawtechUK in the legal sector. This process yields results, but can still be a difficult approach to push sector-wide innovation, as it is primarily based on the energy and availability of these select champions, without a robust data strategy at the organisational level supporting them.

Customer influence on infrastructure

Low 'open culture' maturity within organisations is an issue on the business side of the relationship, as customers, primarily of larger and older organisations, tend to be less interested in, and more resistant to, trying new products and services. It is crucial that the benefits of new services are clearly explained and quantified in order to help large organisations invest internally and build external partnerships.

Research participants from a traditional bank noted that product and service innovation was much easier for challenger and digital-native banks due to their customer base. These banks attract customers who are less risk averse to trialling new financial products, typically away from traditional banks, which gives these challenger banks insight into how to drive customer value through new services.

Firms in the accounting and legal industry are also somewhat dependent on client preferences and market behaviours, which can limit their decisions around technical infrastructure. This can manifest in not wanting to use new solutions, such as aversion to cloud storage and computing, or preferring email over new access protocols. Conversely, when customers are keen to try new technologies, or are undergoing a digital transformation themselves, this affords accounting and legal service providers opportunities to experiment.

Nascent investment in AI/ML proficiency in large organisations

Larger commercial organisations and data providers are only beginning to invest in in-house AI and ML technical proficiency. This is currently happening in dedicated innovation teams as the wider organisation is not yet sure about the tradeoff between risks and benefits. In some organisations this can drive the perception that they are not yet ready for AI/ML solutions.

Technology skills underpinning AI/ML solutions, such as cloud capabilities and data analytics, are present in most large organisations, but have not been mobilised for use in AI and ML. There is a real opportunity to develop these skills across the industry, particularly in medium-sized organisations that tend to be lacking these capabilities, as there has not been a significant client push factor.

In the absence of in-house skills, both large and medium-sized organisations may outsource the needs of AI/ML solutions on a case-by-case basis. These organisations may also enter innovation challenges, or trial different collaborations in order to gain access to these capabilities and help understand their value.

There is also a need for non-technical ML skills such as those related to data and AI literacy, ethics and governance. Large organisations, especially in the finance sector, have been investing in this space through hiring third parties to help perform these roles and train staff, as well as using available tools and methodologies to improve their own practices and principles.

Process insights from market research

Privacy, confidentiality and data management processes

Most organisations across banking, accounting and legal need to follow an extensive system of compliance and control to move data or provide access rights, often with multiple gatekeepers in every step that may differ across organisations. This occurs between firms, but also within firms crossing multiple jurisdictions or within firms spanning multiple sectors, such as accounting and advisory. Companies spanning multiple jurisdictions and sectors are looking for ways to improve the legal sharing of insights and information across jurisdictions without moving data.

Organisations have multiple roles to carry out different data responsibilities. For example, a data protection officer (DPO) in a banking organisation would typically be responsible for data flow within the organisation, while data access needs are met by the information assurance team and customer 'know your client' (KYC) data is managed by the compliance team.

In the legal sector confidentiality is paramount and seen as highly restrictive to data flows and innovation. In our roundtable co-hosted with LawtechUK, 'owing confidentiality to clients' was ranked the top risk to data access and sharing. Despite this, many participants intimated that the catch-all of confidentiality was not always a helpful term and can often be used to avoid seeking challenging yet rewarding solutions. Participants agreed that there would be value in reviewing and stratifying confidentiality within the industry in order to make data and information more accessible, and derive benefits such as operational efficiencies, new services and cost savings from it.

'One-size-fits-all' access models

Organisations in regulated sectors tend towards a one-size-fits-all data access approach based on the strictest privacy and security requirements in the organisation. It is often the simplest approach to complying with regulation in a complex data landscape, with high penalties for mistakes. This over-compliance is often seen as an excuse to avoid unwanted actions.

Many stakeholders believe that varying data access policies based on sensitivity of the data and privacy requirements could improve efficiency in data innovation internally and externally. However, individuals leading efforts to adapt the access model often lack the buy-in and resources to create a powerful case while contending with a strongly-mandated compliance department.

Participants mentioned that when this pain point is voiced at industry-wide forums and in peer networks, it is often met with widespread acknowledgment as a common issue. Coordination at these events, which lack a specific purpose or commitments from attendees aligned with this topic, often does not progress much further than relationship building.

Expanding and overlapping peer networks

Many research participants discussed how they as individuals or their organisations were likely in multiple mission-driven peer networks and industry organisations, in order to seek benefits from greater innovative collaboration. Networks such as Innovate Finance, Tech Nation and DFK International brought together innovation champions and provided a forum for them to engage in greater collaboration. However as mentioned above, these collaborations often lacked significant organisational support.

Participants still appreciated the benefits of these industry bodies, and especially their overlap with the 'official' government-licensed networks such as ICAEW, The Law Society and UK Finance. They believe these linked networks provide opportunities for beneficial partnerships that could result in improved data collaboration and innovation.

Technology-forward departments and partnerships

Proof points in more digitally savvy parts of the business, or innovation trial partnerships are being used to make the case for more investment in digital skills, processes and technologies. Sometimes this produces clashes between traditional IT functions against innovative teams, but other times, champions within those traditional IT functions will use the success of internal experiments or third-party partnerships to make their own cases internally.

Industry bodies and regulators also collaborate with innovative startups in different ways. Some through incubation programmes, such as ICAEW helping grow, regulate and later invest in for-profit company Engine B to provide greater interoperability in the audit and accounting sector. Some through funding research, such as FCA becoming a soft collaborator in RegulAItion's AIR platform, and helping develop a use case for the market.

Banks involved in the open banking ecosystem typically partner with startups to deliver new solutions that drive benefits for their customers. Participants from a traditional bank discussed the idea of being able to start a challenger bank connected to the traditional bank. Such an initiative would allow for greater freedom in experimentation and could be a 'proof point' for the traditional bank and its more risk averse customers.

Innovation programme and challenge popularity

Innovation programmes are popular amongst regulators, data providers and data intermediaries as they provide a good vehicle to experiment with new services, build partnerships and build topic awareness.

They often provide low risk environments for innovation as the data is usually either synthetic or compliant with regulations. Investments in the programmes and challenges are also lower risk investments, due to the relationship building, coaching and marketing that comes with it.

As an additional benefit, organisations, especially data intermediaries, from predominantly different sectors often participate. This allows for a cross-pollination of ideas, techniques and technologies from different sectors.

Non-technical standards for data

In our research on open standards for data, we note that besides technical standards for data such as file formats and data models, non-technical standards for data are important to consider as well. These non-technical standards include codes of practice and importantly, terminology.

Organisations across banking, accounting and legal use different terms such as 'data sharing', 'data access', 'data transfer', and rarely consistently. Broadly speaking, data sharing is perceived as something done externally and involves data leaving a boundary, whereas data access is more internal and data is seen as not moving.

In many cases, including most day-to-day discussion, inconsistent terminology is not an issue. However in heavily regulated areas where legal and regulatory compliance is foregrounded, terminology can break down more open data processes due to fears

that misunderstanding will lead to a legal or compliance breach. Some organisations are trying to mitigate this uncertainty by creating publicly available glossaries, being clear about their language to smooth out these processes.

Technology insights from market research

Low adoption of AI/ML technologies

As reflected above regarding AI/ML skills, organisations have had a low adoption of AI tools and ML algorithms.

'The Kalifa Review of Fintech' echoed the sentiments of many research participants, that 'deep expertise in artificial intelligence, data and machine learning are in short supply within the UK.'[28] Plans to address this include retraining and upskilling adults to meet the needs of UK fintech by ensuring access to short courses from high-quality education providers at low cost.

In the meantime however, large firms are starting to invest more in the digital and data capabilities needed for proper AI/ML adoption. This includes data science departments, and ensuring that AI ethics is an internal priority.

Experimenting with privacy-enhancing techniques

In order to test and implement new data innovations, organisations use a variety of anonymisation and privacy-enhancing techniques on testing and modelling data. In general, organisations are open to multiple PETs but focus on those that make it easiest to share while retaining quality and privacy.

Banking organisations and regulators have primarily used synthetic datasets for machine learning modelling and trialling, preferring this method for ease of sharing and good fidelity overall. Federated learning algorithms are also being deployed to help develop use cases focussed on regulatory compliance and combating financial crime.

Use cases being developed in the accounting and legal sectors have so far focussed primarily on developing federated learning solutions for use cases involving unquoted company valuations, and decisions on legal advice respectively.

Use of third-party providers (TPPs)

Accounting and law firms use TPPs to manage data externally and internally, using the default audit functionalities and inheriting the access rules accordingly. There is an appetite for the use of external partners to provide data management services.

An underlying reason for this is the prevalence of legacy systems that create blockages and bottlenecks for lots of internal development. Participants noted that legacy systems are more likely to get updated when there is external pressure, generally from a business case or partnership involving data access and sharing.

[28] Gov.uk (2021), "The Kalifa Review of UK FinTech", https://www.gov.uk/government/publications/the-kalifa-review-of-uk-fintech

Open banking partnerships have also shown the value of TPPs in delivering innovative new services outside of traditional bank core offerings. These partnerships span from price comparison and micropayment services, to personal finance managers that use increasingly sophisticated algorithms to improve customer financial health.

Sandboxes and scaleboxes

Innovation teams within regulators and large data providers are often unable to get access to as much data as they want in order to run innovation experiments with external partners. In order to work around this barrier they have implemented sandbox technology to provide a safe environment for developers to work on new propositions.

Both the FCA Digital Sandbox Pilot and the Open Banking Directory Sandbox have been cited as examples of regulator-provided sandboxes that provided useful experiences to their partners that they would not have been able to access otherwise. The success of the FCA pilot has led to an extension of the programme both in length and scope. The FCA has also agreed to introduce a scalebox called a ‘regulatory nursery’ in order to help scale-ups with growth and securing investment.

Recommendations

Our research has shown that improving access to data generally, and through innovative privacy preserving means specifically, is anticipated to bring significant benefits to the banking, accounting and legal sectors.

We see that data providers are interested in AI/ML solutions for data-driven regulation, compliance and supervision, and that these solutions are being developed both in-house and by data intermediaries looking to fill gaps in the market, and move it forward. Equally, government, regulators and industry bodies are encouraging innovation in this space, proving the lie that innovation and regulation are at odds with each other, or need to be balanced against each other.

To further explore this area, we have identified a number of recommendations for different stakeholders within the ecosystem, including data intermediaries and data providers, and governments, regulators and industry bodies. Unlocking the benefits of more open and trustworthy data ecosystems will require collaboration across a range of stakeholders.

For data intermediaries

- Build partnerships with regulators, industry bodies and data providers. This will help your organisation demonstrate trustworthiness in the sector, improve sector-specific knowledge, and help develop common standards (such as data models, sector vocabulary and PET best practice), which will drive better data access services.
- Participate in innovation challenges, ideally sponsored by potential partners. This will help to test new product ideas, build valuable connections in the ecosystem and demonstrate the value of product offerings.
- Run sustainable data access workshops to understand what the best mix of revenue streams is, specifically looking at subscription fees, membership fees, and additional or premium services. Open source tools like the ODI's Sustainable Data Access Workbook can help facilitate these workshops.
- Run trustworthy data access workshops to understand what the main signifiers of trust and trustworthiness are in relevant sectors, such as certification, membership of industry networks and commitments to openness, such as publishing open access research and open data. Open source tools like the ODI's Trustworthy Data Stewardship Guidebook can help facilitate these workshops.
- Produce research and case studies around sector and technology trends. This will help your organisation to understand the latest relevant developments in your sector, deliver valuable insights to others in the ecosystem and improve your trustworthiness to potential clients and partners.

For data providers

- Educate users on the new technologies and mindsets required to really thrive in data economies. Utilising modern technologies to achieve better outcomes does not happen overnight, it requires organisational and cultural change. Demonstrate how a mindset shift to secure, privacy-preserving data access will help to overcome common innovation obstacles and empower teams to deliver real change, fast.
- Review current approaches to sharing data, both for regulatory compliance and for commercial propositions. Explore where new approaches to privacy-preserving data access and processes like ML and AI could deliver additional business value. Address internal concerns about the risks of data sharing, and ensure there are appropriate policies that enable safe and automated access to data at scale. Understand where a one-size-fits-all approach is reducing data access and use in less sensitive areas of the business.
- Convene an ecosystem of partners, suppliers and customers to explore:
    - Use cases where improving access to data can create business benefits
    - How to make regulatory compliance more efficient and trustworthy through data-driven processes
    - Foregrounding data ethics within typical business practices
    - Where common standards, from technology decisions, to data model, to vocabulary (such as around confidentiality, or data access vs sharing) could improve data access and interoperability
    - Share best practices and success stories to drive a more open culture in the banking, accounting and legal sectors
- Develop a data strategy that will help to improve data governance, develop the internal data literacy of the business, and promote partnerships with innovative data intermediaries. Locate and encourage innovation champions within the organisation that can lead on initiatives within the strategy. Set measurable objectives to ensure compliance with these goals.
- Explore the commercial landscape for PETs and AI/ML-driven data access and innovation. This can be done through individual partnerships with data intermediaries, running innovation programmes with startups and scaleups, and funding research and development with academia.
- Develop an explicit approach to data ethics and AI ethics. Open source tools like the ODI Data Ethics Canvas and Consequence Scanning can help illuminate current gaps and opportunities.

For government, regulators and industry bodies

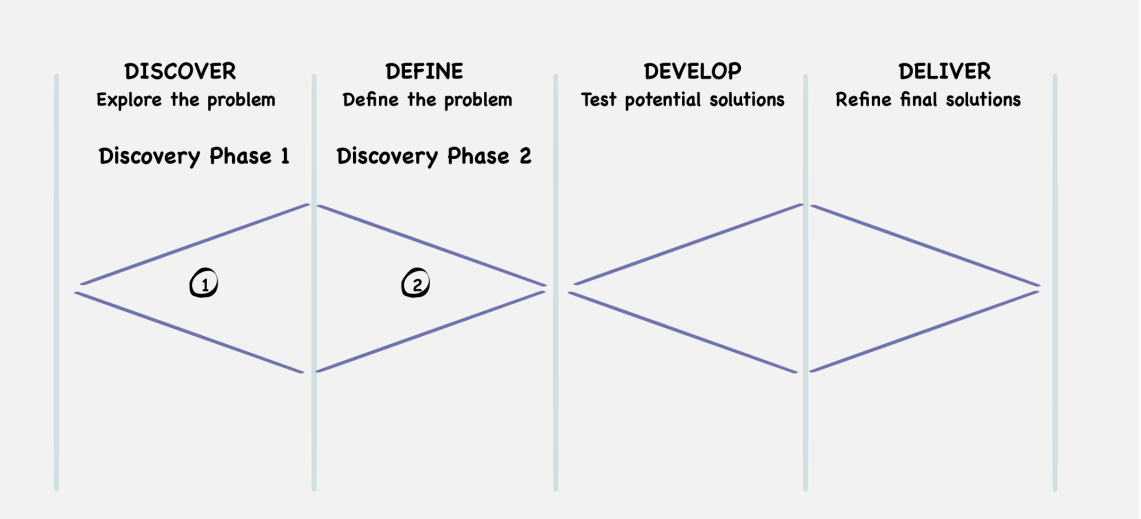
- Convene businesses, including data providers, data intermediaries and other technology providers within their networks to help map out current and future privacy preserving data access arrangements. These could include reusable data sharing frameworks, sector data institutions, and more openly accessible metadata on data that cannot be easily shared.

- Run innovation programmes, like the Digital sandbox pilot: FCA DataSprint, that will help to facilitate open innovation around data access, PETs, and AI/ML product development within specific sectors. Data availability mechanisms should be built into the design and functioning of all such sandboxes as an enduring regulatory principle.
- Fund and support work to scope and incubate businesses aiming to develop PETs to improve regulated data access, that will help organisations within a sector to share data in ways that will preserve privacy and increase trust
- Fund research and case studies around the opportunities for PETs and AI/ML to improve data access in regulated sectors
- Support the development of legal frameworks, guidance, codes of practice and training that will help support ethical, legal and trustworthy sharing of data across a range of sectors. Foreground the development of both technical and non-technical data skills.

Appendix 1: Methodology

We took a holistic approach to gathering the existing thought process of industries involved in the problem space by doing an in-depth exploratory study. Formative research methodologies were used with divergent (phase 1) and convergent (phase 2) thinking.

Phase 1 was about exploring the problem and phase 2 was about defining the problem and figuring out the concrete actions needed to tackle the usage of data.



Our original goal was to interview organisations from the banking, accounting, legal and insurance sectors. Given the unavailability or unwillingness of stakeholders from the insurance industry to participate in our research, the insurance sector was descoped to maintain a robustness of results.

Phase 1/Pilot (Explore the problem)

In this phase, we aimed to start understanding the experience of each participant in order to identify the range of problems that exist. Desk research (literature review) was carried out to get a better understanding of the context and semi-structured in-depth interviews were conducted to understand user pain points, barriers and motivations. Later, insights were drawn out through using thematic coding data analysis techniques and organised by People, Process and Technology.

Why was an in-depth semi-structured interview used as data collection?

The primary sources of data collection in formative research are often semi-structured in-depth interviews. The use of in-depth interviews enables the researcher to explore

issues with participants through encouraging depth and rigour, which facilitates the emergence of new concepts/issues and contributes to the ‘richness of data’ required in the discovery stage. A conversational approach to the interviews was adopted to get deeper context from the participants.

Phase 2 (Define the problem)

From phase 1, we established a range of insights. We prioritised them to focus on and explore a few high-value areas that would help define the key directions, uncover the high-level user needs and identify the riskiest assumptions. To do this we facilitated roundtable discussions, surveys and structured in-depth interviews.

In phase 1, we observed that the participants were wary about speaking on the topics of data access and sharing, due to internal compliance and the impact it could have on reputation. To mitigate the risk of being unable to gather organisations for roundtables and ensure the study was not compromised, the team designed an in-depth interview format and survey as a back-up tool to gather any additional insights.

In general in both interviews and roundtables we employed the Chatham House rule or simply promised that our findings will not be attributed to individuals or organisations. This allowed a more open and timely discussion, as many of the organisations we spoke to have very strict and arduous compliance processes for publicly attributable statements. Engaging in this process would have reduced the value and the volume of the insights we gathered.

In addition we sought to design an effective questionnaire for phase 2 by avoiding biases with questions that were formulated with careful interventions, such as using probes rather than leading questions. However our results in the questionnaire were too low to use independently in our final report, and the findings from it have only been included as part of overall trends in the findings from our research. RegulAItion intends to deploy the questionnaire in future market research where appropriate.

Research questions

The aim of the research study was to understand the potential of, and barriers to, data access in the regulated sectors of banking, accounting, legal and insurance in a way that could both help key stakeholders in those sectors improve data access and to help understand the viability of RegulAItion's AIR Platform proposition.

The research questions that followed were:

1. What is the biggest challenge at the moment in the [name of the industry] industry on how data is shared/accessed?
2. Can you please list the use cases/business needs that require information to be shared?
3. Can you please list the internal process that needs to be taken before information can be shared? Please be specific about each department.

4. How are interactions with datasets tracked and audited to reduce the risk of data misuse?
5. Is the [name of the industry] industry in some way change-resistance?
6. How/where is the [name of the industry] industry investing resources in data processes, people and technologies?
7. How is the industry keeping up with the latest technological advancement especially around ML/AI? Do you think this could be improved? How?
8. What are common misconceptions people have around ML/AI/Data handling? How can we combat these misconceptions and communicate more effectively?
9. How do you communicate as an organisation around 'data access and sharing'? What is the common terminology you use to handle data?
10. What are the challenges around data transfer and management? What are some of the common processes used industry-wide to comply with compliance and control?
11. What does a trusted secure data management platform look like?
12. How are you enabling people in your organisation to help build digital trust?/ How are you engaging your organisation's business processes to build digital trust?
13. What are the specific security challenges in your industry?
14. What are the data sharing risks that your industry needs to be prepared for?
15. What are some of the common data management tools/processes that are in place in your industry and what are the key challenges?

Exploring these questions provided a landscape view of the role of data in the regulated sectors.

Appendix 2: Additional Resources

ODI resources

- The Trustworthy Data Stewardship Guidebook – created to help organisations assess, build and demonstrate trust and trustworthiness
- The Sustainable Data Access Workbook – supports organisations that steward data to make better decisions about revenue models
- 'Data sharing in the private sector' – includes case studies on businesses gaining value from increasing access to data they hold and a summary report
- The ODI's Data Toolkit for Business – a set of collaborative tools that will help businesses unlock the value of data
- 'Anonymisation and open data: An introduction to managing the risk of re-identification' – introductory guide to considering risks around re-identification
- Open Standards for Data – a guidebook that helps people and organisations create, develop and adopt open standards for data
- The Value of Data report – exploring ideas around how to effectively and ethically tap into the value of data
- ODI work with partner Refinitiv exploring the ethical implications of alternative data, and how data intermediaries can build trust in alt data ecosystems
- ODI blog and report on how far Open Banking has come one year on
- The Data Landscape Playbook, which helps organisations involved in data access initiatives – collaborative programmes that focus on addressing social, environmental or economic challenges by improving access to data – to prioritise and plan their initial activities to design or strengthen data infrastructure