
Using data to take an open approach to investment banking

Libby Young, Amanda Smith, Simon Troup
Open Data Institute supported by Deutsche Bank

4	About	
5	Executive summary	
	Draw on industry expertise with an independent facilitator	5
	Manage control and access of data	5
	Assess opportunities for a more open approach that creates long-term value	5
6	About the investment banking sector	
7	What is the challenge?	
9	Defining data infrastructure for investment banking	
	What data assets are produced in investment banking?	9
	What technologies and processes are used in investment banking?	10
	What organisations are involved in investment banking?	11
12	Building a more open data infrastructure	
	The Open Banking Standard	12
	The Open Protocol	12
13	Recommendations	
	Draw out industry expertise via working groups	13
	Manage control and access of data	13
	Assess opportunities to create long-term value	13
14	Open approaches to data problems: existing policies and technologies	
	Collaborative maintenance models	14
	Case study: Legislation.gov.uk	14
	Case study: OpenStreetMap	14
	Using open data to map beneficial ownership	14
	Case study: OpenCorporates	15
	Open standards for data	15
	Open registers and APIs	15
	Open source code and shared vocabularies	15
	Blockchain technology	16
	Case study: Everledger	16
	Case study: Provenance	16
17	Conclusion	
	Annex: The ODI's design principles for data infrastructure	17
	Bibliography	18

About

This report was produced by the Open Data Institute and supported by Deutsche Bank.

About the ODI

The ODI works to build a strong, fair and sustainable data economy by helping governments and businesses around the world get data to people who need it. It is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data.

Authors: Libby Young, Amanda Smith, Simon Troup

Editors: Anna Scott and Charlotte Fleming

Design and art direction: Adrian Philpott / PHILPOTT design



Executive summary

Investment banking is one of the world's oldest and most data-rich industries. It provides a large range of financial products and services to businesses, governments and investors, helping them to grow and to manage risk. The sector is undergoing significant change: adopting new technologies and regulations, transforming organisational cultures and reducing costs.

Investment banks collect, share and use data every day to make decisions and execute transactions.

The return on data held by banks has historically been limited by a low interest in data infrastructure amongst client-facing teams and friction in data processing.

However, as clients want more data – and as technology advances and costs less – a more open approach is emerging. At the same time, regulators are mandating banks to share more, and the growth of open data in other sectors has shown its positive impact on innovation. A more open data infrastructure in investment banking is inevitable, and banks must respond and adapt in a timely and informed manner.

The ODI wants to support this shift by facilitating a discussion on investment banking data infrastructure, and working with stakeholders to identify, test and implement solutions that make data more accessible. This report considers:

- the sector's existing data infrastructure – in terms of data assets, processes, technologies and organisations
- where data assets are currently mapped on the Data Spectrum, from closed to open
- challenges and opportunities in using open data to create a strong data economy
- open data case studies from other industries

To take an open approach to the investment banking sector's data infrastructure, we make the following recommendations.

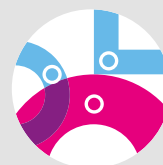
Draw out industry expertise via working groups

- Working groups can help bring industry expertise into an agile decision-making framework. Working groups could harness collective insights from banks into where the greatest data challenges lie.



Manage control and access of data

- Organisations can be protective of the data they control, because of the competitive advantages they think it affords them. However many groups have a stake in banking data – clients, competitors and wider ecosystems. Open solutions can provide more flexible ways to manage how data is controlled and accessed.



Assess opportunities for a more open approach that creates long-term value

- An open solution that serves the needs of all key stakeholders ensures committed and aligned participation from across the industry. Some organisations committing time or capital may not experience an immediate return on investment from a more open data infrastructure, but sustainable data initiatives will ultimately benefit all industry participants.



The report then highlights existing policies and technologies that the investment banking sector can take inspiration from in finding open approaches to data problems: collaborative maintenance models, mapping beneficial ownership, open standards, open registers and APIs, open source code and shared vocabularies and blockchain technology.

About the investment banking sector

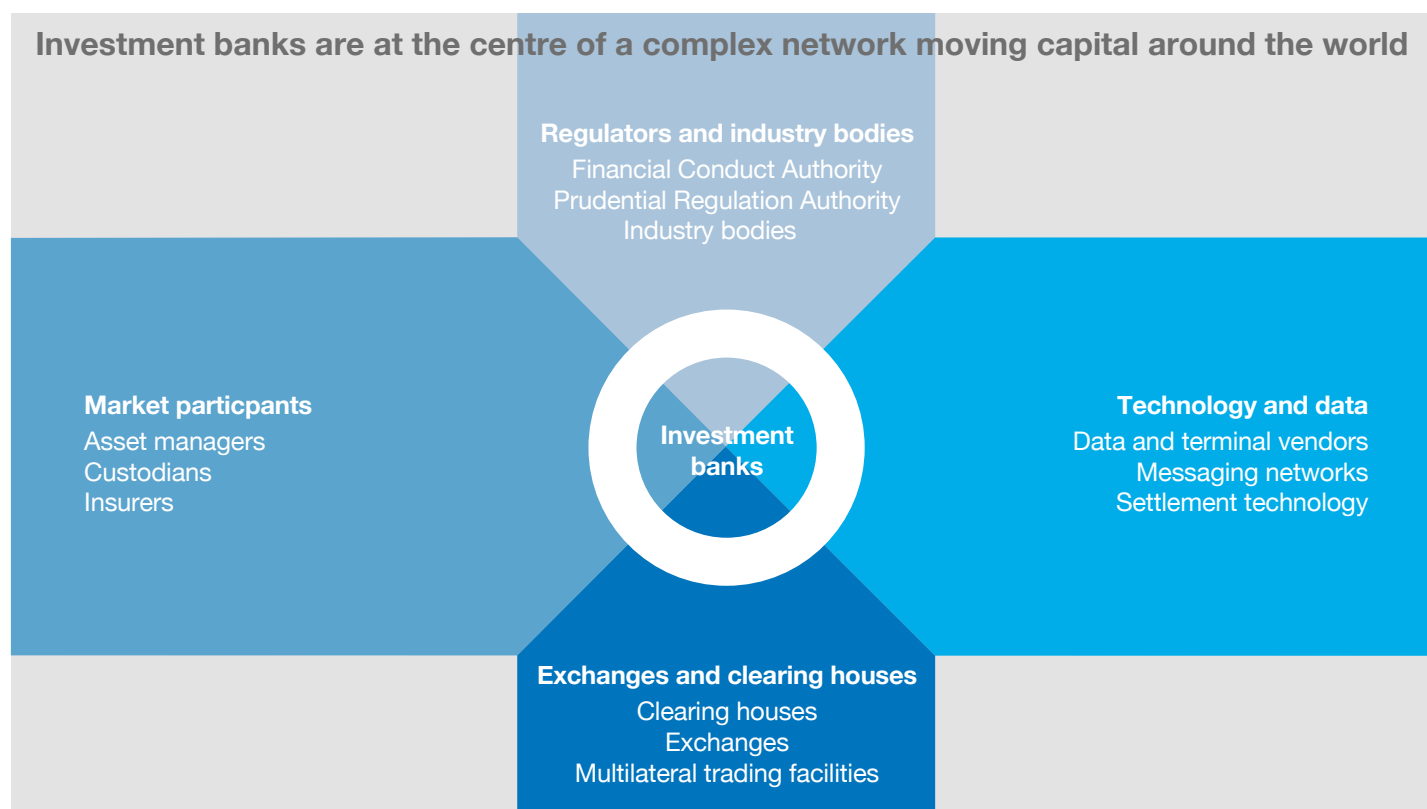


Figure 1: Investment banks are at the centre of a complex network, moving capital around the world.

The role of investment banking has changed over the years.

Investment banks emerged in the 1400s to support merchants trading in goods such as silks, metals and spices. In these early manifestations, trust and risk management were important measures for those using these services. Since then, the role of investment banking has changed, and the sector has continued to adapt at the centre of a complex network of institutions.

1 Bank for International Settlements (2012), 'Principles for financial market infrastructures', <http://www.bis.org/cpmi/publ/d101.htm>.

2 L Michael Meyer (2016), 'Regtech 123', <https://www.linkedin.com/pulse/regtech-1-2-3-l-michael-meyer-cfa?trk=hp-feed-article-title-like>.

Regulators, industry bodies and investment banks are exchanging vast amounts of public and private data, while the latter facilitates the movement of investment capital around the world. In this modern manifestation, data is particularly valuable to building trust and understanding risk.

In recent years, efforts by regulators to improve transparency and stability^[1] in the sector have significantly increased costs for investment banks,^[2] as they struggle to keep up with demand for better ways to track and secure data that is exchanged. Low interest rates and advances in technology have also put pressure on traditional operating models.

At the same time, the speed of technology and infrastructure innovation in capital markets – where investment banks primarily operate – has been slower than in retail and corporate banking.

What is the challenge?

Data is underused in investment banking.

Investment banks, their clients and regulators, increasingly view data as an underused strategic asset and are actively exploring new ways to maximise its value and potential.

Three factors in particular are driving this shift:

- the industry's growing recognition of shared needs that it can collectively address and benefit from
- clients and regulators requiring more and better data, and clients in particular looking for ways to access data that match advances in consumer technology
- innovation in other industries being driven by more open infrastructures

The sector collects, shares and uses vast quantities of data every day, primarily via transactions. In 2015, the value of shares traded was \$99.8 trillion^[3] (greater than global GDP), and the harder-to-track derivatives market is estimated to be at least 20 times that.

Operationally, administrative data and data-related tasks alone are estimated to generate \$4.4bn annually in IT and processing costs.^[4] And while 70% of investment banks know data quality affects costs, only 11% actively measure the cost of bad data quality.^[5] Finally, 80% of any bank's data is unstructured and dynamic, which the sector is only in its early stages of exploring how to use.

What do we mean by static and dynamic data?

'Static' data is a fixed dataset that does not change once it is created. 'Dynamic' data changes as further information becomes available.

3 The World Bank (2017), 'Stocks traded, total value'

<http://data.worldbank.org/indicator/CM.MKT.TRAD.CD>.

4 BCG Perspectives (2016), 'Fintech in Capital Markets: A Land of Opportunity' <https://www.bcg.com/en-gb/publications/2016/financial-institutions-technology-digital-fintech-capital-markets.aspx>.

5 Accenture (2016), 'Reference Data Management: Understanding True Cost', https://www.accenture.com/t20151202T165846__w_/us-en/_acnmedia/Accenture/next-gen/top-ten-challenges/challenge3/pdfs/Accenture-2016-Top-10-Challenges-03-Reference-Data.pdf.

6 See: <https://permid.org>.

7 Thomson Reuters (2017), 'KYC onboarding still a pain point for financial institutions', <https://blogs.thomsonreuters.com/financial-risk/know-your-customer/kyc-onboarding-still-a-pain-point-for-financial-institutions>.

In researching this paper, sector participants highlighted three broad data challenges in investment banking:

Search: discoverability and accessibility of data

Trust: transparency of data provenance

Quality: accuracy and timeliness of data

At the organisational level, the relevance of each data challenge varies across different parts of a bank. Bespoke, lower frequency investment banking services (top-left in Figure 2, overleaf) rely more on **search** and **trust** – here, data has a material impact on reputational, counterparty or market risk. Higher-volume and more commoditised businesses (bottom-right) tend to focus on data **quality** and cost – they use large amounts of data where errors have a smaller economic impact per transaction, but in aggregate are material.

The following three examples demonstrate the breadth of impact that more open and shared data could have in the sector.

- 1) Historically **proprietary reference data** – for example, Bloomberg's Instrument Reference Data or Reuters Instrument Codes (RIC) – add costs and processes to data access and collection, as well as the potential for errors. A more open infrastructure could remove these barriers, something which is already being explored by Thomson Reuters in the form of PermID.^[6]
- 2) **Know-Your-Customer (KYC) processes and regulations** – which guide identity management and help to establish data provenance – can be a burden for banks and clients alike. A global survey carried out by Thomson Reuters found that the average onboarding had reached 26 days, and half of those respondents thought the time taken to onboard would increase in the coming year.^[7] Furthermore, annual KYC costs for large intermediaries can be several hundred million dollars. Although a handful of utility-like KYC entities selling due diligence data services have emerged in recent years, KYC remains cumbersome and costly. A more collaborative, sector-wide process with the appropriate permissions could simplify counterparty risk management, improve customer experience, and aggregate data to develop new client services.

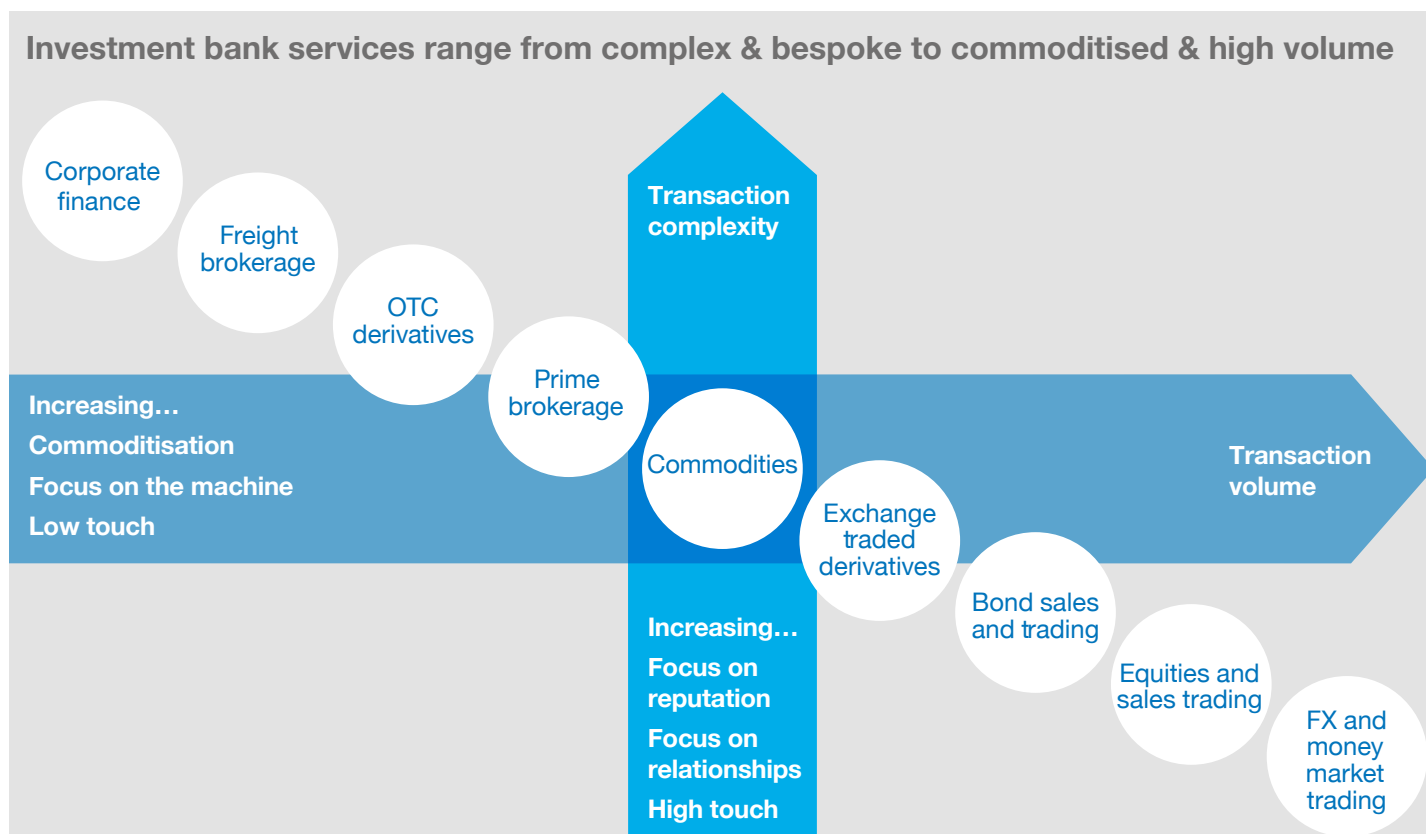


Figure 2: Investment bank services range from complex & bespoke to commoditised & high-volume.

- 1) The **Markets in Financial Instruments Directive (MiFID II)**, comes into effect in Europe in January 2018. The directive aims to enhance market efficiency and resilience and will impact the entire trade lifecycle and all asset classes. It also increases requirements for real-time data and effectively mandates institutions to develop more open data infrastructure internally. If banks agree standards for sending data to Data Reporting Services Providers (DRSPs), over time those standards could be used to communicate directly with each other, speeding up execution and settlement processes, improving counterparty risk management capabilities, and possibly leading to new services.

Many institutions manage their data in silos, with unnecessary duplication and poor integration across different business lines and functions. An internal reorganisation of data is a major challenge, but could also reveal ways these newly aggregated datasets could be appropriately repurposed and shared to advance the sector's core services.

Defining data infrastructure for investment banking

Each sector is underpinned by data infrastructure – datasets, technologies and processes, along with the organisations that maintain and govern them.

Data sits on a spectrum, falling between closed data (typically used internally within institutions), shared data (accessible to groups across institutions and sometimes paid for), and open data (available for anyone to access, use and share).^[8] The more open and accessible an industry's data infrastructure, the more trust and value it tends to create.^[9]

What data assets are produced in investment banking?

The highly transactional nature of investment banking creates vast quantities of data daily that constantly need to be stored, accessed and updated. There are many ways to categorise this data, but it can be broken down into five broad

categories: contracts, identities and securities, which are typically more static (i.e. do not change once created); and ledgers and transactions, which are typically more dynamic (i.e. change as further information becomes available).

Each of these categories of data can be mapped to the Data Spectrum.^[10] Client contracts would represent 'closed' data and be kept private, while identities are 'shared' within or between contracting organisations. Ledgers are 'shared' with regulators and shareholders, and transactions reported by exchanges and securities information are made public, although not necessarily 'open', for anyone to access, use and share.

There are many opportunities to address challenges and create business value by moving data towards the more open end of the spectrum. Simply including a maintenance API in an individual bank's securities system could extend the community of data contributors within the organisation beyond a central team. This captures benefits of open networks, such as getting more perspectives scrutinising data to improve its quality and empowering a more diverse group of users to fix it.

Data category	Example/s from the investment banking sector
Contracts	Counterparty agreements for derivatives transactions
Identities	Client identities, Know Your Client (KYC), risk profiles
Ledgers	Accounts (exposures, positions, custody records)
Transactions	Orders, executions, confirmations, settlements (MiFID)
Securities	Instrument identifiers (Sedol, Ticker), Pricing, Initial Public Offering (IPOs), Bonds

8 The Open Data Institute (2017), 'The Data Spectrum', theodi.org/data-spectrum.

9 The Open Data Institute (2017), 'Principles for strengthening our data infrastructure' theodi.org/guides/principles-for-strengthening-our-data-infrastructure.

10 The Open Data Institute (2017), 'The Data Spectrum', theodi.org/data-spectrum.

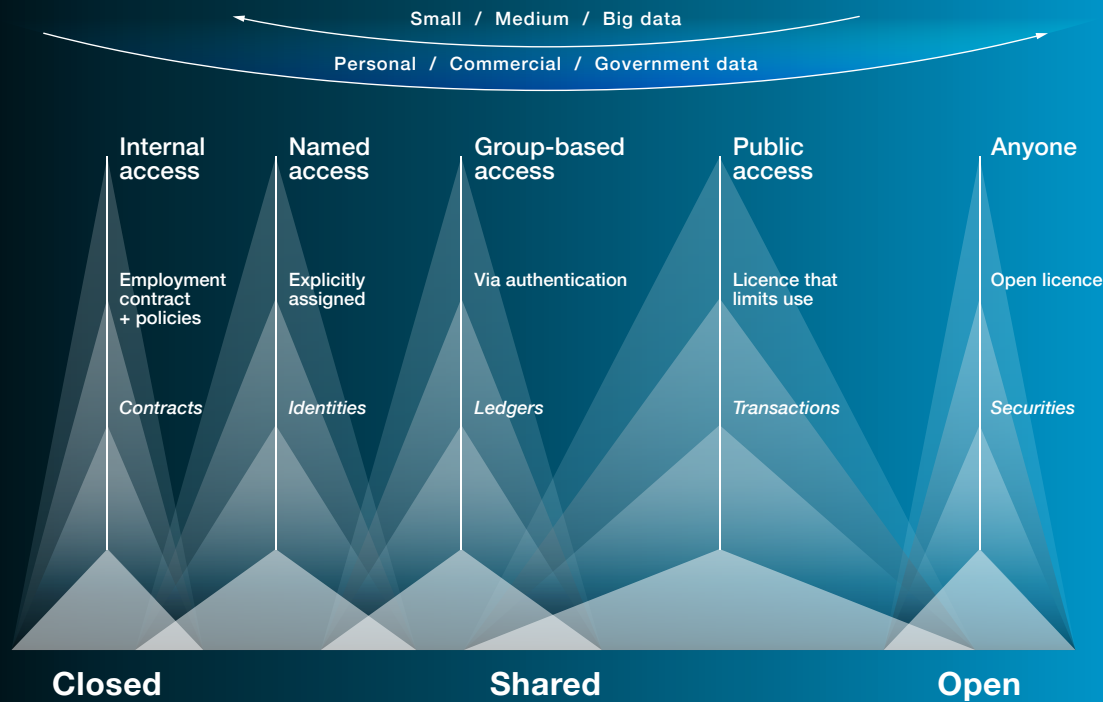


Figure 3: The Data Spectrum that the ODI proposes for the investment banking sector.

What does 'open' mean for investment banking?

Open data is data that anyone can access, use or share. An 'open' application programming interface (API) does not automatically make the data it is delivering open, rather its technology and the standard itself. The data could be made accessible by being 'shared' or made 'open'. Shared data would only be accessed with appropriate permission (whether from an individual or business), and subject to the API's approved security and technical standards.

However, it is not appropriate to make all data open: data should be kept closed or shared where it is necessary to protect personal or commercial privacy. That said, where data is kept closed or shared, it should not be because of habit, or a lack of knowledge about the possibilities created by open and accessible data. It is important to make deliberate decisions about how accessible data should be or where it should be on the spectrum.

What technologies and processes are used in investment banking?

After the 'Big Bang' in 1986, when financial markets were deregulated and the London Stock Exchange was reformed, the sector saw a move from traditional face-to-face share dealing to electronic markets. This brought a drive towards 'straight through processing' (STP) to optimise transaction processing speeds, using tools and standards that were new at the time such as Extensible Markup Language (XML) and large-scale messaging infrastructure. Benefits included ridding the process of paper-based trading and creating electronic markets.

Investment banking has become increasingly commoditised with a greater focus on centralised data services and a platform approach. Underpinning banking technology are data standards and shared vocabularies, such as data naming and definition, interfaces including real-time and intermittent interactions, and technical data infrastructure, such as databases, cloud and SaaS. Banks have become increasingly 'joined up' with core investment activities; finance, risk and compliance functions, serviced by centralised technology teams, for example.

What organisations are involved in investment banking?

As intermediaries, investment banks sit at the centre of a complex network of institutions and data flows. Market participants, exchanges and clearing houses collectively deliver a robust infrastructure of capital markets, overseen by regulators and supported by technology and data services.

Only through engagement with all stakeholders can banking infrastructure become more open, transparent and robust, or deliver a lasting trust. While regulators such as the Prudential Regulation Authority (PRA)^[11] protect the public from systemic risk, understanding the network of transactions and connected entities is essential. When a regulator stress-tests the market for the failure of a bank, they need to understand:

- the transactions that bank is engaged in, such as the exchange of financial instruments between parties (or ‘swaps’), which occur on a daily basis
- who holds debt that may default
- which hedge funds hold collateral that could be lost in an insolvency

Two trends worth noting here are the innovative regulatory sandboxes being championed in the UK^[12] and Australia,^[13] and the industry’s increasing commitments to accelerators and collaborative partnerships with new fintech^[14] entrants. Both these developments are indicative of a more collaborative and innovative culture and bode well for the development of an open data infrastructure in investment banking.

11 Bank of England (2017), ‘Prudential Regulation Authority’, <http://www.bankofengland.co.uk/pr/Pages/default.aspx>.

12 Financial Conduct Authority (2015), ‘Regulatory Sandbox’, <https://www.fca.org.uk/firms/regulatory-sandbox>.

13 Australian Securities & Investment Commission (2017), ‘Innovation Hub’, <http://asic.gov.au/for-business/your-business/innovation-hub/>.

14 BCG Perspectives (2016), ‘Fintech in Capital Markets: A Land of Opportunity’, <https://www.bcgperspectives.com/content/articles/financial-institutions-technology-digital-fintech-in-capital-markets/#chapter1>.

Building a more open data infrastructure

Strengthening data infrastructure and making it as open as possible presents opportunities for investment banking.

The Open Banking Standard

In the UK, the CMA-endorsed Open Banking Standard for retail and corporate banking offers an example of an open approach being applied to improve efficiency and stimulate innovation.^[15]

The Open Banking Standard was developed by the Open Banking Working Group (OBWG) to guide how open banking data should be created, shared and used.

New legislation – the Revised Payment Service Directive – will come into effect in January 2018, mandating banks within the EU to offer open application programming interfaces (APIs) to licensed third-parties, who can provide account or payment services to customers. The directive is intended to improve innovation and give customers more choice.

This prompted the OBWG to be set up, at the request of HM Treasury, to explore how data could be used to help people to transact, save, borrow, lend and invest their money.

Making it possible to share data that banks have historically held can improve people's banking experience. For example, when data is securely shared or published openly using open APIs, it can be used to build useful applications and resources to help people find what they need. Customers can look for a mortgage more easily, banks can find customers matched to a new product, and businesses can share data with their accountants. This, in turn, will improve efficiency and stimulate innovation.

The OBWG collaboratively developed an Open Banking Standard to guide how open banking data should be created,

shared and used by its owners and those who access it.^[16] This is being carried out by an Open Banking Implementation Entity.

While PSD2 undoubtedly increases costs and competition for incumbent banks in the short-term, those who have embraced the open standard in the UK are starting to demonstrate its value – banks are giving developers access to their APIs, and their Open Up Challenge has been designed to drive innovation for their customers and themselves.^[17]

The Open Protocol

The Open Protocol is an open standard created for the hedge fund industry.^[18] It standardises how hedge funds collect, collate and convey risk data, making aggregated risk reporting (ledgers) more accessible and accurate for clients and regulators. The data is not centrally held, and hedge fund managers have complete control over the data and how it should be distributed.^[19]

The implementation process from inception to launch and adoption took nine months. This speed and impact was achieved due to:

- a common need and purpose shared by hedge funds, investors and regulators
- the absence of a commercial agenda for the initiative itself (with a clear path to value-generation for all participants)
- the leadership of an independent working group using an open and federated approach
- a collaborative and iterative process, with input from industry bodies and regulators observing the collaboration, and refinement by public consultation

The solution is simple and pragmatic: templates give investors access to standardised Value at Risk (VAR) numbers by asset class, sector and region, and shows explicit properties of the risk methodology being applied. By using standard templates to collect information, the industry gets a flexible and transferable framework for risk aggregation and information exchange. An accompanying manual sets out the processes to populate the templates to produce a standard output.

The Hedge Fund Standards Board (HFSB) has become the co-chair of the Open Protocol Working Group and has been added to the Standards Body for Alternative Instruments (SBAI) toolbox.^[20] The Open Protocol template is currently used by funds with over \$1 trillion in assets under management.^[21]

15 The Open Data Institute (2015), 'The Open Banking Standard', <http://theodi.org/open-banking-standard>.

16 Open Banking (2017) 'About Open Banking', <https://www.openbanking.org.uk/about>.

17 NESTA (2017), 'Open Up Challenge', <http://www.nesta.org.uk/project/open-challenge>.

18 The Open Protocol (2017), 'Open Protocol', <http://www.theopenprotocol.org>.

19 Ibid.

20 Standards Board for Alternative Investments (2017), 'Toolbox', <http://www.sbai.org/toolbox>.

21 The Open Protocol (2017), 'Open Protocol', <http://www.theopenprotocol.org>.

Recommendations

We recommend three ways to make data infrastructure for investment banking stronger and more open.

Draw out industry expertise via working groups

Working groups can help bring industry expertise into an agile decision-making framework. For the investment banking sector, working groups could harness collective insights from banks into where the greatest data challenges lie.

Setting up these groups can be difficult, with coordinating organisations needing to engage many market participants. Having one or two organisations in a trusted, independent position to convene and facilitate is key.

For example, the Open Protocol began as a joint effort between regulators and the online hedge fund community The Albourne Village,^[22] both well-aligned with the hedge fund industry. The Open Banking Standard was a joint effort between the UK government and the retail banking sector, facilitated by the ODI.

Manage control and access of data

Organisations can be protective of the data they control, because of the competitive advantages they think it affords them. However many groups have a stake in banking data – clients, competitors, regulators and wider ecosystems. Open solutions can provide more flexible ways to manage how data is controlled and accessed. We encourage investment banks to review our proposed Data Spectrum for investment banking to see how it aligns with their existing data asset registers and flows.

The Open Protocol's methodology for describing risk allows asset managers and investors to retain a level of control over data storage and distribution. The Open Banking Standard gives retail bank customers more control over data about them, and the ability to decide who they share data with.

Assess opportunities for a more open approach that creates long-term value

An open solution that serves the needs of all key stakeholders ensures committed and aligned participation from across the industry. Some organisations committing time or capital may not experience an immediate return on investment from a more open data infrastructure, but sustainable data initiatives will ultimately benefit all industry participants.

The Open Protocol demonstrates how fast and effective implementation can be when key industry stakeholders all have a clear rationale for collaborating – hedge fund managers, institutional investors and regulators all benefit from its structured risk framework.

One area where there may be a clear and immediate need is regulation. Another area where there is a clear rationale for an open approach, but perhaps less impetus for change, is the overall trade life cycle.

By using these three techniques, the investment banking sector would be able to understand the value of collaborating on a more open data infrastructure, which could lead to specific datasets and processes to be developed and opened up.

The ODI's principles for data infrastructure provide a framework against which to apply these three techniques. See the Annex for more.

22 Albourne Village (2017), 'The Albourne Village', <https://village-eu.albourne.com>.

Open approaches to data problems: existing policies and technologies

Patterns are emerging in how policies and technologies are being used to solve data problems across sectors.

With the investment banking sector prime to explore better ways to maximise data's value and potential, it can be useful to consider policies and technologies for collecting, sharing and using data that have been developed to achieve the same aim across different sectors.

Collaborative maintenance models

Collaborative maintenance models help mitigate the cost of managing and updating data across a community of users, while improving data discoverability and quality, improving audit and control capabilities, and facilitating engagement and innovation. The legislation.gov.uk and OpenStreetMap case studies below show how such a model can work.

Case study: Legislation.gov.uk

Historically, it has been hard to maintain legislation in accessible and up-to-date formats. The National Archives' [Legislation.gov.uk](http://legislation.gov.uk) overcame this challenge by using an open API and a collaborative maintenance model.^[23] Their innovative 'expert participation programme' includes participants from public and private sector groups who all saw the value that open, up-to-date legislation would bring them. As a result, in the last few years over half of the updates to legislation have been made by these experts, five times faster than before and bringing 80% of all legislation^[24] on the website up to date.

Case study: OpenStreetMap

OpenStreetMap has demonstrated how collaborative maintenance is an extremely efficient way of managing rapidly evolving data.^[25] Following the 2010 Haiti earthquake, emergency services lacked an up-to-date map of roads and the newly established resources and camps that supported relief efforts. With the help of satellite imagery and volunteers, NGOs were equipped with digital maps within days.^[26] OpenStreetMap empowered users to take control of data about them, improving quality and making content more reactive to change.

In investment banking, testing and implementing this approach could begin with technically simple models such as registers^[27] (explained further in the 'Open registers and APIs' section) and focus on necessary but non-competitive datasets.

For this, the investment banking sector should seek to explore:

- how sector registers might work with a more open model
- if register members could take a more direct role in maintaining data about themselves
- what data governance processes would be needed to make that work
- what value a more open model would create

Using open data to map beneficial ownership

The securities lending market has become a particular focus in regulators' efforts to promote market stability and enhance corporate transparency. Open data can better track the true beneficial owners of securities to prevent fraud, money laundering and corruption.^[28] They can also help ensure proxy votes contributing to decisions on management pay awards, and ensure other topics are more transparent and accountable.

23 See: <http://www.legislation.gov.uk>.

24 The Open Data Institute (2017), 'Case study: Legislation.gov.uk', <https://theodi.org/case-studies/case-study-legislationgovuk>.

25 See: <https://www.openstreetmap.org/about>.

26 Wikipedia (2017), 'OpenStreetMap', https://en.wikipedia.org/wiki/OpenStreetMap#Humanitarian_aid.

27 Government Digital Service (2015), 'Registers: authoritative lists you can trust', <https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust/>.

28 Open Ownership (2017), 'About the project', <http://openownership.org/about>.

Case study: OpenCorporates

OpenCorporates is the largest database of companies in the world, with data from 116 million firms in 119 jurisdictions.^[29] Its founders saw that open company data from central governments was valuable but hard to search for and lacked structure. Data was available for those with the resources to collect and structure it, but there was no product fulfilling that need at scale and in an open and flexible way. Coinciding with the evolution of graph databases for modelling social networks, OpenCorporates built a fast and flexible data structure to make search and visualisation fast and pain-free. It removed the complexity of multiple data sources, delivering easily searchable, open content and broadening access to this data. This supported entrepreneurs, innovators and those with fewer resources.

Open standards for data

Open standards are created through an open process which ensures they take into account the requirements of multiple stakeholders. Having different organisations adopting the same open standards ensures that data can flow easily between them, and that they share the same tools.

If open standards for data were used in investment banking, they could support processes or products, which would reduce friction in how data is exchanged, helping to improve client onboarding, increase transparency and save costs.^[30]

Regulatory technology (RegTech) solutions seeking to reduce costs and simplify processes have emerged in recent years, but open standards have yet to be tried in investment banking. Open standards could prove most effective for the highest impact and largest-scale challenges. For example, KYC could use them to support identity management and help establish data provenance.

From a process perspective, banks have sought to make end-to-end transactions as 'light touch' as possible, in order to increase returns. The absence of open standards causes most friction when data is sent between banks or even internally between business functions, as data is often missing, inconsistent or hard to transfer.

In public markets that use exchanges – such as equity or stock markets – the underlying cause of friction is usually data quality (i.e. a company's stock ticker is freely available, but entered incorrectly at one stage in a process).

29 See: <https://opencorporates.com>.

30 Thomson Reuters (2016), 'Thomson Reuters 2016 Know Your Customer Surveys Reveal Escalating Costs and Complexity', <https://www.thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html>.

31 OpenSource.com (2017), 'What is open source?', <https://opensource.com/resources/what-open-source>.

32 See: <https://openaps.org>.

33 See: <https://www.gnucash.org>.

34 W3C (2014), 'Best practices for publishing linked data', <https://www.w3.org/TR/ld-bp>.

35 International Swaps and Derivatives Association (2016) 'New Industry Standard for Calculating Initial Margin Widely Adopted by Market Participants', <http://www2.isda.org/news/isda-simm-deployed-today-new-industry-standard-for-calculating-initial-margin-widely-adopted-by-market-participants>.

In private markets, such as OTC derivatives, friction is usually caused by discoverability and accessibility of data, particularly around contracts and identifiers. It is hard to find or transfer the data attached to a new derivatives transaction. If banks collectively solved data discoverability issues in private markets, it would create opportunities to transfer these learnings into a new model for public markets.

Open registers and APIs

A register is an authoritative list of data that can be trusted. An open register can move beyond relying on datasets that are updated periodically and may have errors, to operating on data that is trustworthy, standardised and open for scrutiny and improvement. An open API is an application programming interface that provides access for all to a web service or software application. It does not necessarily make the data it is delivering open, but makes the technology and the standard itself open.

Open registers and APIs for data that is not commercially sensitive (but is required to execute transactions) could provide significant opportunities for investment banking. A simplified and coordinated standard for country, bank and account codes could be accessed by open registers and/or open APIs and reduce the number of steps required to execute an international transfer, thereby reducing the number of transfer failures caused by data issues.

Open source code and shared vocabularies

Open source code can be inspected, modified and enhanced by anyone.^[31] Projects that use open source code include the Open Artificial Pancreas System (APS) project, which is an open and transparent effort to make safe and effective basic APS technology widely available to reduce the burden of Type 1 diabetes.^[32] Open source code is also adopted by GnuCash, which provides free accounting solutions for personal and small business accounts.^[33]

Shared (or standardised) vocabularies help facilitate inclusion and interoperability, for example, the Data Catalog Vocabulary (DCAT) is a vocabulary designed to facilitate interoperability between data catalogs published on the web. By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs.^[34]

Investment banks have long been users of open source code in their back office operations, but embedding it into interactions with clients, regulators and shareholders offers opportunities to create value.

Some investment banks have begun giving clients access to open source technology, and open source code and data ontologies are emerging, particularly in reporting. The open business reporting language by extensible business reporting language (XBRL) can be used for both regulatory and shareholder reporting. OpenGamma helps derivatives market participants implement new International Swaps and Derivatives Association (ISDA) standards to calculate how much capital must be put aside for certain types of trades.^[35]

Opportunities remain to harden regulation and reduce the burden on investment banks by tightening and linking data standards (e.g. MiFiD). Equally, wherever regulation and standards are in question, the debate about how best to balance accuracy and transparency continues.

Blockchain technology

Blockchain technologies are distributed ledgers that provide a way to store information so that many people can see it, keep a copy of it, and add to it. Once added, it is very difficult to remove information, which reinforces trust in blockchain content.^[36] Essentially a blockchain is a database shared across the web with many people holding a copy of it, which shows a single version of the truth.

Everledger's work with diamonds demonstrates how fraud can be reduced as physical assets are certified and tracked digitally. Provenance's work shows how blockchain technology can help makers tell the stories behind their products to customers; it is used by over 200 retailers and producers in the food and drinks industry.

Case study: Everledger

Everledger use blockchain, smart contracts and other emerging technology to support banks and insurers to reduce risk and fraud.^[37] They do this by improving supply chain transparency and efficiency. Everledger believe a reduction in document tampering leads to a reduction in fraud for industries.^[38]

Everledger started by applying blockchain technology to the diamond industry. After creating relationships with the major certificate houses around the world, they added data about diamonds to a blockchain and have now uploaded data about 1,000,000 diamonds.

Everledger are using this secure and transparent record to track and protect these valuable assets by recording all transactions throughout the supply chain without relying on intermediaries. It helps people to understand the provenance of a diamond, who bought it and owns it, if it has been resold and where it is now. By adding transparency to the diamond supply chain, Everledger hope to use blockchain technology to fix a market full of corruption, trafficking and violence. Everledger work with Ebay to track the resale of diamonds, Interpol to tackle crime, insurance companies to tackle fraud and banks to assist in double financing of diamonds. They see their approach being applied to other high-value goods, from other gemstones through to wine, art and watches.

Case study: Provenance

Recognising increased public awareness of product origins and history – such as the sustainability of materials, conditions of workers, and carbon footprint – Provenance provided a platform to build a trustworthy supply chain ledger. It improves transparency by unmasking opaque chains of data, helping customers make informed purchasing decisions.

Provenance uses blockchain technology to help build trust in goods and their supply chain, and uses open data from partners including Sourcemap.com and OpenCorporates.^[39]

Over 200 retailers and producers in the food and drinks industry use their software service to help prove the provenance of their product.^[40] For example, the Co-op food and digital teams use the software to track produce from source to shelf in real-time.^[41] Their work with Fair Food verifies proof of payment for living wage to 55 farmers while tracking coconuts from South East Asia to Europe.^[42] As well as restoring connections between makers, sellers and customers, Provenance brings value to smaller makers that cannot become transparent on their own due to the cost of opening their data and representing it on the web in a meaningful way to customers.^[43]

Blockchain technologies could disrupt existing processes in investment banking by delivering transformative change to how banks store and exchange data. Near instantaneous settlement of trades could become a reality when asset ownership is entirely digitised in this fashion. As distributed data technologies evolve, we may see more edge-based computing that facilitates secure and trustworthy open data products across the banking network.

36 The Open Data Institute (2016), 'Applying blockchain technology in global data infrastructure', <https://theodi.org/technical-report-blockchain-technology-in-global-data-infrastructure>.

37 See: <http://everledger.io>.

38 YouTube (2016), 'Everledger & Diamonds: Building a Secure Blockchain', <https://www.youtube.com/watch?v=sRVwkzQi5hI>.

39 See: <https://www.provenance.org>.

40 Ibid.

41 See: <https://www.provenance.org/case-studies/co-op>.

42 See: <https://www.provenance.org/case-studies/fairfood>.

43 See: <https://theodi.org/case-studies/provenance-case-study>.

Conclusion

A more open data infrastructure presents opportunities for investment banking across the value chain.

In the last decade, the significant increase in financial regulation has in part been a reaction to the financial crisis but also a recognition of longer-term trends: the exponential growth of capital markets, the vast amounts of structured and unstructured data created by it, and changes in the infrastructure to support it.

Taking a more open approach would help regulations to be implemented in ways that drive interoperability and innovation, for example making it easier to collect and verify data. Open solutions and open approaches to data for the sector are most likely to emerge where they benefit three key stakeholder groups: investment banks, their clients and regulators.

Regulation is a good place to start, so long as participants can see a clear path from protocol enhancements to creating value. The most significant opportunities often lie in the largest but least transparent areas, such as unstructured data and derivatives markets.

Perhaps the most critical factor in introducing a more open data infrastructure will be the process used to implement it. As the Open Banking Standard and the Open Protocol show, using an independent facilitator to harness industry expertise can be a fast and simple way of producing results. Equally, designing a process with no commercial agenda itself, but that delivers long-term value for participants, removes potential barriers to cooperation amongst competitors. In building a more open data infrastructure, the biggest shift for investment banks will be to agree where collaboration creates more value than competition.

At a systemic level, a more open infrastructure for investment banking will support distribution of capital globally and more effective risk management. From the ambitious entrepreneur to the individual seeking affordable health insurance, such a system has the potential to benefit many.

44 The Open Data Institute (2016), 'We need to learn how to search the web of data', <http://theodi.org/blog/we-need-to-learn-how-to-search-the-web-of-data>.

45 Lateral Economics (2016), 'Permission granted: The economic value of data assets under alternative policy regimes', <http://theodi.org/research-economic-value-open-paid-data>.

Annex: The ODI's design principles for data infrastructure

1. Design for open

Open data, open culture, open standards, open source and collaborative models build trust, reduce cost and create more value than other approaches. Being open improves quality as more people can contribute to the outcome, and it increases the number of connections that can be made. Data benefits from network effects: it creates more value as more people use, contribute to and maintain it.

2. Build with the web

We need to learn how to publish, discover, use and link together data across the web.^[44] Data on and in the web is continuing to grow with more devices being connected and interconnected every day. The billions of people, sensors and services on the web produce and use data. Data infrastructure must support the web of data.

3. Respect privacy

In the most impactful and valuable data infrastructure openness is maximised, but what is private remains private. Different countries have their own data protection legislation and social contracts, which need to be adhered to. To build trust, organisations using personal data should also be open with people about how they use and share that data.

4. Benefit everyone

Data infrastructure components should be designed and supported to benefit as many stakeholders as may use it. Everyone should benefit from the innovation, services and insights that the whole data infrastructure allows. Sometimes data infrastructure that is as open as possible will benefit the organisations that maintain the data, in other cases it will not.^[45] To benefit everyone, it will be necessary for governments to provide support for some components.

5. Think big but start small

Don't start big. Start with the problems that are making it hard for people to make decisions or build new services, be agile and learn from experiments. Concrete and tar don't go out of date as quickly as data technologies do.

6. Design to adapt

Expect needs to change, and expect other needs to vary between different stakeholders and local contexts. Be prepared to experiment with new technologies and ideas, look for desire paths, measure impact, learn from what works and what doesn't. Any part of data infrastructure might start as a small experiment but turn out to create significant impact and have high demand. If it is designed to adapt using approaches like human-centred design, by encouraging innovation and by using flexible modular approaches, this is most likely to happen.

7. Encourage open innovation

The best ideas can come from anywhere: individual citizens, large or small organisations and from the public, private or third sectors. Strong data infrastructure and open innovation will encourage and stimulate fair and equitable markets and innovative ways to both maintain data and use it to create new services.

Bibliography

Accenture (2016), 'Reference Data Management: Understanding True Cost', https://www.accenture.com/t20151202T165846__w_/us-en/_acnmedia/Accenture/next-gen/top-ten-challenges/challenge3/pdfs/Accenture-2016-Top-10-Challenges-03-Reference-Data.pdf

Australian Securities & Investment Commission (2017), 'Innovation Hub', <http://asic.gov.au/for-business/your-business/innovation-hub>

Bank for International Settlements (2012), 'Principles for financial market infrastructures', <http://www.bis.org/cpmi/publ/d101.htm>

Bank of England (2017), 'Prudential Regulation Authority', <http://www.bankofengland.co.uk/pr/Pages/default.aspx>

BCG Perspectives (2016), 'Fintech in Capital Markets: A Land of Opportunity', <https://www.bcg.com/en-gb/publications/2016/financial-institutions-technology-digital-fintech-capital-markets.aspx>

Financial Conduct Authority (2015), 'Regulatory Sandbox', <https://www.fca.org.uk/firms/regulatory-sandbox>

Government Digital Service (2015), 'Registers: authoritative lists you can trust', <https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust>

L Michael Meyer (2016), 'Regtech 123', <https://www.linkedin.com/pulse/regtech-1-2-3-l-michael-meyer-cfa?trk=hp-feed-article-title-like>

Lateral Economics (2016), 'Permission granted: The economic value of data assets under alternative policy regimes', <http://theodi.org/research-economic-value-open-paid-data>

International Swaps and Derivatives Association (2016), 'New Industry Standard for Calculating Initial Margin Widely Adopted by Market Participants', <http://www2.isda.org/news/isda-simm-deployed-today-new-industry-standard-for-calculating-initial-margin-widely-adopted-by-market-participants>

NESTA (2017), 'Open Up Challenge', <http://www.nesta.org.uk/project/open-challenge>

Open Banking (2017), 'About Open Banking', <https://www.openbanking.org.uk/about>

OpenSource.com (2017), 'What is open source?', <https://opensource.com/resources/what-open-source>

Standards Board for Alternative Investments (2017), 'Toolbox', <http://www.sbai.org/toolbox>

The Open Data Institute (2016), 'Applying blockchain technology in global data infrastructure', <https://theodi.org/technical-report-blockchain-technology-in-global-data-infrastructure>

The Open Data Institute (2017), 'Case study: Legislation.gov.uk', <https://theodi.org/case-studies/case-study-legislationgovuk>

The Open Data Institute (2017), 'The Data Spectrum', <http://theodi.org/data-spectrum>

The Open Data Institute (2015), 'The Open Banking Standard', <http://theodi.org/open-banking-standard>

The Open Data Institute (2017), 'Principles for strengthening our data infrastructure', <http://theodi.org/guides/principles-for-strengthening-our-data-infrastructure>

The Open Data Institute (2016), 'We need to learn how to search the web of data', <http://theodi.org/blog/we-need-to-learn-how-to-search-the-web-of-data>

The Open Protocol (2017), 'Open Protocol', <http://www.theopenprotocol.org>

Thomson Reuters (2017), 'KYC onboarding still a pain point for financial institutions', <https://blogs.thomsonreuters.com/financial-risk/know-your-customer/kyc-onboarding-still-a-pain-point-for-financial-institutions>

Thomson Reuters (2016), 'Thomson Reuters 2016 Know Your Customer Surveys Reveal Escalating Costs and Complexity', <https://www.thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html>

W3C (2014), 'Best practices for publishing linked data', <https://www.w3.org/TR/ld-bp>

Get in touch

theodi.org | [@ODIHQ](https://twitter.com/ODIHQ)

info@theodi.org

Open Data Institute, 65 Clifton Street, London EC2A 4JE

