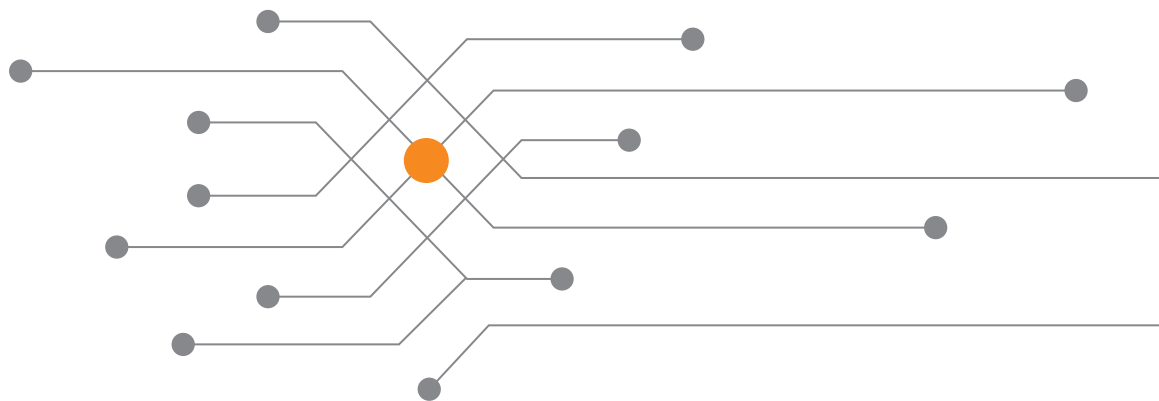
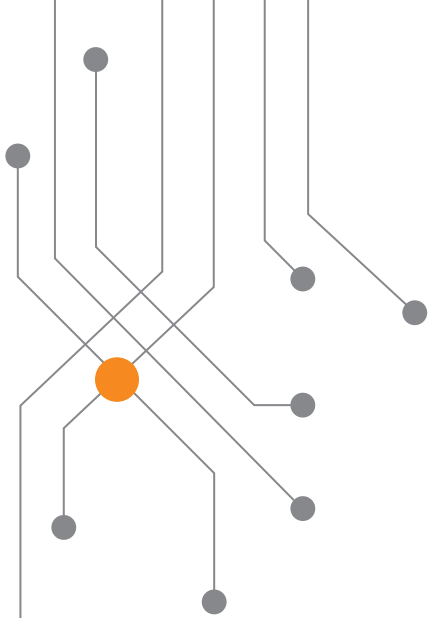


STRENGTHENING OUR LEGAL DATA INFRASTRUCTURE



Contents

INTRODUCTION	5	CASE STUDY 2: COURT LISTINGS	16
MANAGEMENT SUMMARY	6	Data assets	16
THE ROLE OF DATA IN THE LEGAL LANDSCAPE	7	Balancing accessibility, transparency and privacy	17
The importance of data in the legal sector	8	CASE STUDY 3: CASE TRANSCRIPTION DATA	17
Overview: design principles for data infrastructure	9	Data assets	17
Understanding the data landscape	10	Opinion piece: deploying technology to speak truth to justice	18
Making wider use of data	10	CASE STUDY 4: LEGISLATION	19
Opinion piece: why Europe needs a more strategic approach to Legal Open Data	12	The impact of Brexit on legislation	20
Overview: business models for data in legal	13	CASE STUDY 5: BRINGING OTHER DATA ASSETS TO THE LEGAL PROFESSION	21
CASE STUDY 1: CASE LAW AND RELEVANT DATA FROM THE COURTS	13	Land Registry	21
Data assets	13	How is Land Registry data being used?	21
Technologies and processes	14	Companies House	22
Analogue origins dictating digital data flows	14	How is Companies House data being used?	22
Interpretation for different users	14	CONCLUSION	23
Opinion piece: freemium platforms can play a major role in the Open Law Movement	15	Acknowledgements	23
Opinion piece: Nick West, Chief Strategy Officer, Mischon de Reya LLP	16		



This paper focuses on four central case studies of the legal sector's data infrastructure: case law; courts listings; transcripts; and, legislation. Other data assets that are highly relevant for the legal sector, such as the Land Registry and Companies House, are also discussed as a fifth case study.

Introduction

The global legal system is primarily an information system. At a fundamental level, the information contained in a country's statute law and regulations describes the decisions that its successive governments, elected or otherwise, have made about how its people and organisations can act. It is the "instruction manual for running a country".

In addition, in a common law country like the United Kingdom, the information contained in judicial decisions describes customary rules of behaviour that have evolved over centuries to produce the sprawling system of precedents that influence its citizens' daily lives.

Because this information is highly complex, tens of thousands of people train for the legal profession in order to be able to understand and apply it. As well as learning the law itself, professionals must master yet another type of information: the procedural; behavioural and cultural codes that determine how lawyers interact with the courts; their clients; and, each other.

The legal system is therefore based on a huge volume of written information. The aim of this paper, jointly produced by Thomson Reuters and the Open Data Institute, is to look at this information not as a collection of documents, but as data. If we apply principles of good data infrastructure, what lessons can be learned to aid lawyers, citizens and businesses?

In the past, information and data from statutes, regulators, the courts system and government in general has been locked away in paper documents. There is a wealth of data, such as transcripts or legislation, which people may think of as documents, rather than data assets. Today, while in some instances it is available online, its utility is limited by difficulties discovering the data, access restrictions and costs, as well as its incompleteness and lack of richness.

In this paper, we view this information and data as infrastructure underpinning the legal system and enabling it to operate effectively. We examine what it will take to strengthen the UK's legal data infrastructure, looking at lessons and recommendations, and what this means for the many organisations and individuals who interact with that infrastructure.

This paper focuses on four central case studies of the legal sector's data infrastructure: case law; courts listings; transcripts; and, legislation. Other data assets that are highly relevant for the legal sector, such as from Land Registry and Companies House, are also discussed as a fifth case study.

Whilst a significant portion of this paper focuses on data within the UK, the lessons here are relevant for other jurisdictions. The opinion pieces, overviews of key topics, and developments in the paper show insights from other countries that can inform the creation of a strong legal data infrastructure anywhere.

Public and private sector organisations, communities and citizens, have a shared interest in strengthening our legal data infrastructure. However, as the UK navigates its way through Brexit, improving access to the right legal data will become even more important, as changes to policy, legislation and court hierarchies create rapid changes to law. Thomson Reuters and the Open Data Institute hope that this paper will inspire continuing discussion on this topic of vital interest to national wellbeing.

We would like to acknowledge and thank all those within the Open Data Institute, Thomson Reuters and other organisations who contributed to and reviewed the material herein.

Signed,

Lucinda J Case

J. Tennison



LUCINDA CASE
Lead, Legal Professionals
Europe, Thomson Reuters



JENI TENNISON OBE
CEO, Open Data Institute

Management summary

The legal system is based on a huge volume of written information. In this paper we have set out to think of this as data—and in turn how applying the principles of good data infrastructure can aid lawyers, citizens and businesses.

Data infrastructure is one foundational pillar that is needed for a strong, fair and sustainable data economy. Public and private sector organisations, communities and citizens have a shared interest in strengthening their legal data infrastructure.

In the opening section of this paper on the legal landscape we look in more detail at:

- **The importance of data in the legal sector:** Whilst data from the legal system, such as courts listings and legislation, is now more available than it has ever been before, **accessibility** remains an issue as it is still difficult to navigate where that data is located and how it is licensed. Lack of good **data infrastructure** creates inequality for users of the legal system.
- **Understanding the data landscape:** The legal sector contributes some £25.7bn p.a. to the UK economy.¹ We discuss how the government should balance accessibility of data with creating opportunities for value to be added on top of the data to facilitate **legal interpretation**.

A better data infrastructure, along with the development of strong **data skills** in and around the legal system, will enable the creation of new private and third sector services to assist citizens and businesses who currently cannot access legal advice.

- **Making wider use of data:** We examine how different open data business models are being explored in the legal sector. Further value in creating a strengthened legal data infrastructure comes from the **interoperability** of data across the data spectrum to realise the true potential value of that data through legal **innovation**.

Whilst this paper does not set out a blueprint for the future infrastructure of the legal system, we have used a series of case studies to explore in depth different parts of the legal data landscape—such as case law, court listings, case transcripts, legislation and other data sets of value to the legal system.

The legal sector contributes significantly to the global economy. The Brexit process in the UK highlights that access to the right legal data will become even more important, but this need is global and extends far beyond Brexit. Doing business and living in a globally connected, changing world means that it is vital that we strengthen our legal data infrastructure.

1. www.lawsociety.org.uk/support-services/research-trends/a-25-billion-legal-sector-supports-a-healthy-economy

The role of data in the legal landscape

The legal landscape presents fantastically vast quantities of data. In England and Wales alone, each year millions of court users interact with a system comprised of 77 Crown Court centres, more than 170 county courts, more than 330 magistrates' courts, more than 40 district registries of the High Court, and more than 60 types of tribunal, plus the Court of Appeal and Supreme Court.² In terms of legislation, each year Westminster produces dozens of Acts of Parliament, while Government creates thousands of Statutory Instruments.

In addition, a plethora of Government departments, agencies and regulators and independent bodies generate a huge volume of policies, codes, data and decisions that are communicated and shared by a variety of means, with various implications for citizens, businesses and their legal advisers.

The cross-section of people who use or are affected by this data touches all corners of society: law firm partners and associates; legislators and government lawyers; barristers; in-house lawyers; local government lawyers; and judges; through victims, criminal defendants, court reporters and journalists—right down to the person buying a house or inheriting from a will.

For example:

- A lawyer preparing for court depends on access to prior case law to construct their argument
- Civil servants and elected legislators depend on a full understanding of existing legislation in order to draft policy
- Specialist legal reporters and the press rely on accurate and timely court listings to facilitate reporting

The legal data used in these examples is explored in more detail through the case studies in this paper. By making this data infrastructure as open as possible, there are opportunities to unlock more value from legal data. The case studies in this paper describe different data assets; the technologies and processes involved; how data flows; and, how it can be assessed.

As Figure 1 shows, data access sits on a spectrum, ranging from closed to shared to open.³ Data that needs to be kept private should be kept private (closed). Sensitive or commercial data may be shared with some people or organisations. Data that can be open should be open. As legal data

By making this data infrastructure as open as possible, there are opportunities to unlock more value from legal data.

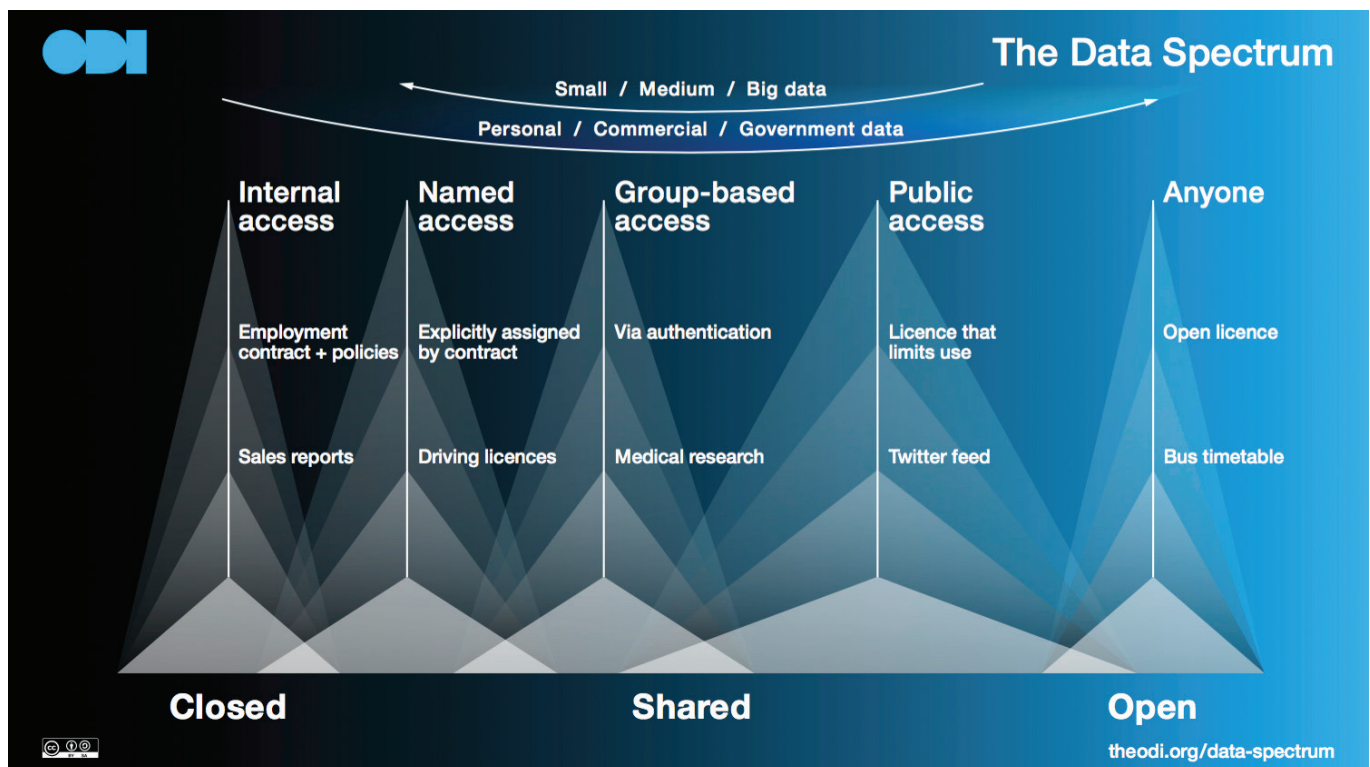


Figure 1. The Data Spectrum, created by the Open Data Institute (ODI)

2. www.judiciary.gov.uk/you-and-the-judiciary/going-to-court

3. theodi.org/data-spectrum

can include personally identifiable information—for example, addresses of defendants in a criminal case, or family courts where entire judgements are sometimes not published to protect the rights of children—it is not appropriate for all legal data to be open.

Furthermore, good-quality data can be easily shared, easily processed, has guaranteed availability and consistency over time, and is traceable so that its users can determine whether or not they can trust the source of the information. It is important that these characteristics are embedded into all data, regardless of whether closed, shared or open. This topic was covered in depth in a previous paper by Thomson Reuters and the ODI, ‘Shareable by Default.’⁴

The importance of data in the legal sector

ACCESSIBILITY

The UK Government’s stated commitment to *Transforming Our Justice System* recognises the importance of technology in modernising the legal sector.⁵ The Lord Chancellor, Lord Chief Justice and Senior President of Tribunals have committed to a significant financial investment, building on recent and ongoing programmes, to bring radical change to the levels of IT in the justice system. One of its key aims is to make sure justice is accessible, with the criminal justice system and tribunals system becoming ‘digital by default’. Under the current plans for HM Online Court, some classes of dispute may eventually be resolved largely by automated technology. However, transforming the sector and unlocking the value of new technologies will require an effective courts data infrastructure.

In parallel with this is the UK’s voluminous legislative environment. Westlaw UK⁶ alone contains around 4,000 Westminster Acts of Parliament and some 60,000 Statutory Instruments. In addition to information from Westminster, there is a growing volume of statute coming from devolved legislative bodies in Scotland, Wales and Northern Ireland. This information is available to varying degrees from a variety of official websites.

Within law firms and legal departments themselves, the trend towards digital technology⁷ is putting

data at the heart of workflows. Matter management software,⁸ document automation, e-billing and other data-rich solutions are seeing widespread adoption by the sector. The popularity of artificial intelligence⁹ platforms like IBM Watson, Ravn and Kira will accelerate this process. It is clear that the data landscape across legal is headed for huge growth and change. Tapping into the collective power of the data landscape requires effectively bridging public and private databases—an invaluable opportunity that is yet to be fully realised.

INFRASTRUCTURE

In this paper we set out to view data in the legal sector through the lens of commonly accepted data infrastructure principles.

Data is critical infrastructure. The volume of data which we create, collect, maintain and store increases each day. Data helps us to make decisions, build services and gain insight. Data infrastructure consists of data assets, the organisations that operate and maintain them and guides describing how to use and manage the data.¹⁰ Data infrastructure underpins business innovation, public services and civil society. It connects different parts of our society and economy and can be strengthened to benefit everyone. The ODI and other organisations are working to define the characteristics of a strong data infrastructure. These include the FAIR (Findability, Accessibility, Interoperability and Reusability) principles, which are recognised by the scientific data and research communities, and the ODI’s design principles for data infrastructure¹¹ and certificates¹². Governance is an important part of data infrastructure, and at present there is little governance for legal data other than for data protection, as led by the Information Commissioner’s Office (ICO).¹³ Research by the British Academy and the Royal Society¹⁴ found that governance frameworks need to ensure trustworthiness and trust in the management and use of data as a whole.

Whilst data from the legal system, such as courts listings and legislation, is now more available than it has ever been before—areas still remain in which the lack of accessible information and data create frustrations and inequality for users of the legal system.

Good-quality data can be easily shared, easily processed, has guaranteed availability and consistency over time, and is traceable so that its users can determine whether or not they can trust the source of the information.

4. innovation.thomsonreuters.com/en/labs/shareable-data.html

5. www.gov.uk/government/uploads/system/uploads/attachment_data/file/553261/joint-vision-statement.pdf

6. legalresearch.westlaw.co.uk

7. blogs.thomsonreuters.com/legal-uk/2017/02/22/legal-it-landscapes-tech

8. A system for managing case files.

9. blogs.thomsonreuters.com/legal-uk/2017/02/20/legal-ai-beginners-guide

10. theodi.org/what-is-data-infrastructure

11. theodi.org/guides/principles-for-strengthening-our-data-infrastructure

12. certificates.theodi.org/en

13. ico.org.uk

14. royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf

Overview: design principles for data infrastructure

Jeni Tennison, CEO, Open Data Institute

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 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.¹⁵ 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. 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; and, 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. This is most likely to happen if it is designed to adapt using approaches like human-centred design, by encouraging innovation and by using flexible modular approaches.

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.

Data benefits from network effects: it creates more value as more people use, contribute to and maintain it.

15. theodi.org/blog/we-need-to-learn-how-to-search-the-web-of-data

It is still difficult to **find** and **access** that data in ways that can unlock value for citizens and businesses. Data could contribute more value to our economy, and to the legal sector, than it currently does. For example, data can create greater value when it is machine readable, and licensed to enable **reuse**.

- Findability, accessibility and reusability are aided by making our data infrastructure as open as possible
- Better data infrastructure can both help improve the current justice system and support transformation
- A more accessible legal data infrastructure can help citizens understand it better

Understanding the data landscape

LEGAL INTERPRETATION

Access to data is not the same as access to justice. More than two thirds (68 percent) of people would not know whether to proceed with a legal matter if they could not talk to a lawyer for advice, and only two in five people (39 percent) think the justice system works well for citizens.¹⁶ Organisations which fight to support access to justice for those not able to engage with the legal system play a vitally important role. Despite prior action, and future intent, to make statutory, judicial and regulatory information more accessible, it still generally requires interpretation to be of use. This is true both for those who can make use of paid legal services and those for whom legal costs are prohibitive and effectively create a barrier to engaging with the legal system.

If government has a clear role in providing access to data then its role in providing interpretation of the law is less clear cut. Government encourages and supports third sector and private sector organisations to provide interpretation of the law and value-added services for various groups of people with their own needs. The services and business models of these private and third sector organisations are underpinned by the data infrastructure that government, courts and others provide.¹⁷

DATA SKILLS

As data becomes a more important part of our legal system, we also need to look at the skills of those that work within the legal system, as well as those whom encounter the legal system.

Although data is increasingly being used by governments, businesses, and people to make decisions, many people still lack the skills to understand how data is being used or how to use data to achieve their goals. The legal process is difficult for outsiders to understand, and while increasing amounts of data are being collected, few of those who are affected by the legal system understand how data is accessed, shared, and used. For example, in the area of court recordings, many litigants may not even be aware of the existence of such recordings, or any potential right they have to access the data about them which has been collected.

At the same time, lawyers, policy makers and those providing services need to be thinking about using data as a tool to achieve their goals. By equipping themselves with the right skills, and working closely with innovators, decision makers can understand where technological change is occurring and where they can benefit from a stronger data infrastructure.

Making wider use of data

INTEROPERABILITY

Most useful applications of data require data from multiple sources to be combined. This is the third paper produced by Thomson Reuters and the Open Data Institute and this principle of interoperability draws on our recommendations from the first two: 'In Creating Value with Identifiers in an Open Data World'¹⁸ we showed the power of having unique, open identifiers; and in 'Shareable by Default'¹⁹ we made recommendations for managing other aspects of data to make it shareable.

16. www.citizensadvice.org.uk/Global/CitizensAdvice/Crime%20and%20Justice%20Publications/Responsivejustice.pdf

17. Access to justice, along with criminology topics such as prison, probation, rehabilitation, restorative justice and others, arose during a workshop that was held to shape this paper. The notes are available for anyone else to use freely when covering those topics, if they wish to use and draw on to make recommendations. <https://theodi.org/blog/strengthening-our-legal-data-infrastructure>

18. innovation.thomsonreuters.com/en/labs/data-identifiers.html

19. innovation.thomsonreuters.com/en/labs/shareable-data.html

INTEROPERABILITY IN THE CRIMINAL JUSTICE SYSTEM

Ongoing efforts by the Crown Prosecution Service and HM Courts and Tribunals Service to establish a Common Platform within the criminal justice system highlight the importance of easily shareable data.

The criminal justice system comprises of a number of agencies, such as the police, the Crown Prosecution Service (CPS), the courts, the National Offender Management Service²¹, and the National Youth Justice Board²². These organisations work together to deliver criminal justice. Users must navigate these, engaging with a number of different agencies throughout the journey, and these agencies have an overlapping network of data exchange.²³ There is no unique identifier for crimes that runs from the police service to the CPS and onwards to the Courts. This makes trying to track a crime through the whole criminal justice system automatically almost impossible, and interoperability of data throughout the system is challenging. A number of problems dogging the system (delays, ineffective trials, lack of coordination) can be traced back to problems with sharing data and information efficiently.

The data spectrum²⁴ is a worthwhile guide when thinking about how all the data that flows through the legal data infrastructure is connected. The topic of interoperability does not imply one joined up legal data 'platform' to consume all that data, but rather the way in which data sources can, and will be combined, in commercial and not-for-profit initiatives in the future.

Interoperability is also about thinking how data flows and connects between these different systems. As government, and business, becomes more engaged to use shared data infrastructure, such as common identifiers, open standards and application programme interface (APIs)²⁵, joining up data will become easier, and the results more powerful.

INNOVATION

The legal sector contributes some £25.7bn p.a. to the UK economy.²⁶ Government and companies can increase this contribution by ensuring the legal sector is underpinned by FAIR data²⁷ to enable new products and services to be developed. Effective data infrastructure will help to create a strong, fair and sustainable legal marketplace.

To make the best use of data, we should look beyond our own teams, organisations and sectors. There needs to be connections between the public sector, private sector, third sector, academia, and a thriving startup ecosystem where new ideas can flourish. The ODI believes that by working collaboratively, and in the open, that we can transform the ways in which we are able to work with data.

- Open data encourages participation and facilitates collaborative methods of collecting and correcting data.
- Open standards and open APIs share and transfer restricted data, accelerating collaboration.
- An open culture encourages participation, debate and sharing knowledge for all.

An ecosystem of startups has already emerged, with innovators tackling risk and compliance, practice management, law for good, analytics and search, contracts and market places.²⁸ The Legal Geek Startup Map²⁹, supported by Thomson Reuters, lists a rapidly increasing number of emerging startups in the UK which are focused on innovation in the legal domain. Examples include Lexoo³⁰, a platform to easily source and compare quotes from specialised commercial lawyers, and CrowdJustice³¹, a crowdfunding platform built for legal cases. Startups have even been formed to help other startups with their legal needs to continue to grow the market. SeedLegals³² enables founders and investors to create all the legal documents they need to build their company and close their funding rounds.

To make the best use of data, we should look beyond our own teams, organisations and sectors. There needs to be connections between the public sector, private sector, third sector, academia, and a thriving startup ecosystem where new ideas can flourish.

21. The National Offender Management System covers prisons and probation.

22. The National Youth Justice Board oversees the Youth Offending Teams.

23. Digital Justice: Criminal Justice Services Landscape - www.flickr.com/photos/gdsteam/20351061738

24. theodi.org/data-spectrum

25. Application Programme Interface (API) is a system of definitions, protocols and tools which enables developers to create software applications that exchange information

26. www.lawsociety.org.uk/news/documents/legal-sector-economic-value-final-march-2016

27. www.force11.org/fairprinciples

28. www.legalgeek.co/startup-map/#LGMap

29. www.legalgeek.co/startup-map

30. www.lexoo.co.uk/how-it-works

31. www.crowdjustice.com

32. seedlegals.com/about-us

Opinion piece: why Europe needs a more strategic approach to Legal Open Data

Marie Bernard, Europe Director of Innovation at Dentons, Strategic Advisor of Nextlaw Labs, and European Legal Tech Association (ELTA) Board Member

The Knowledge Future - Intelligent Policy Choices for Europe 2050 analyses how the Internet has disrupted corporate strategies and the ownership of ideas:

A bewildering range of IPR models has resulted, (...) varying from a collaborative commons to corporate concentration. But the general trend favours consolidation in sectors where multinationals operate. This all puts Europe at a disadvantage: it is generating a smaller share of the world's ideas than in the past, and its influence on global standards – a vital aspect of international trade – has diminished.

As the legal industry witnesses early stages of market consolidation, it's the right moment to reflect on our aspirations for tomorrow's standards.

Europe's current legal data environment can be characterized as follows:

- **Fragmented landscape:** In Europe, by nature, law as big data is fragmented. Each state's legislation, international agreements, and case law—not to mention European and international laws—are stored in different national and supranational databases.
- **Increasingly affordable product development:** Inventing a new product is difficult and expensive, but as technology mainstreams and algorithms are copied, production costs decrease and the barriers to entry are lowered.
- **Augmented legal knowledge:** As indexing, abstracting legal content, and providing metadata are disrupted by artificial intelligence, retrieving and processing legal knowledge will become exponentially faster and more reliable.

Law firms, corporations and legal tech startups are now considering legal big data in their commercial strategies. If Europe's legal industry were to consciously and deliberately co-create common legal data standards, what would that look like?

- **Lobby for the adoption of legal open data policies.** This would support an overall improvement in the quality and consistency of datasets, stimulate reuse initiatives and unlock public-private collaboration. This could benefit not only commercial usage but also broader purposes.

- **Overcome our short-term individual aims and national standards.** Ahead of market consolidation, we need a continental perspective on open legal resources and unique identifiers for multiple sources of law. We may have to crowdsource the setup and maintenance of data sets, metadata and ontologies.
- **Approach this challenge as a consortium.** Instead of fragmented "me too" initiatives, we should come together and seek expert advice on best standards of linked data, semantic web and datasets management. Other industries have paved the way for such an approach.
- **Compete on quality.** Europe could differentiate and compete on the quality and sharpness of the solutions re-using data, offering unique blends of proprietary, paid-for and open data.

A more strategic approach to Legal Open Data in Europe could be a concrete instrument for the legal industry to meaningfully contribute to its own transformation while supporting the communities it serves.

While there is no guarantee that collaborative, bottom-up alternatives will emerge based on legal open standards, the bigger risk may well be to miss out an extraordinary opportunity! Ultimately, open data could bring us closer to achieving the targets of the United Nations' Sustainable Development Goal on Peace, Justice and Strong Institutions:

- Promote the rule of law and ensure equal access to justice for all
- Develop effective, accountable and transparent institutions
- Ensure public access to information and protect fundamental freedoms

A more strategic approach to Legal Open Data in Europe could be a concrete instrument for the legal industry to meaningfully contribute to its own transformation while supporting the communities it serves.

Overview: business models for data in legal

David Curle, Director of Market Intelligence, Thomson Reuters

There are many examples and approaches that businesses are using to engage and embrace closed, shared and open legal data. These examples could underpin a more open legal data infrastructure:

Re-sell a high-quality, integrated data source.

Many publishers take open data and add value in the form of editorial enhancements (annotations, cross-references, summaries, indexing, integration of various sources, etc.). This is perhaps the oldest business model in the legal space; since the 1800s, publishers have taken raw legal data and made it more valuable by adding layers of interpretation, finding tools, and tools for updating research to help lawyers ensure that the law is in full force. This is the approach that has been a cornerstone of the business models of the largest legal publishers including Thomson Reuters, LexisNexis, and Wolters Kluwer.

Provide free access to a subset of data, but charge for a separate, more comprehensive offering.

This is the model applied by Fastcase's Public Library of Law or Thomson Reuters' Findlaw. Each of them are separate offerings that have

a limited set of data. The idea is that revenues will come through upselling to the full service (Fastcase) or through other revenue streams (Findlaw's marketing services offered to law firms).

Provide free search and retrieval of open data as part of a legal research service, but charge for more advanced features and capabilities.

This is the approach of some of the newer US legal research players such as Casetext and Ravel Law. Case law can be retrieved for free, but analytics and other services that leverage and build on the data require a subscription.

Third-party payment through professional affiliations.

Fastcase and Casemaker are low-cost research providers who make their services available to state and local bar associations (for free or at reduced rates) as a benefit of membership.

Free access as part of an advertising and analytics-supported free search system.

This model is applied by Google. Its collections of case law and scientific literature are available in Google Scholar.

Case study 1: case law and relevant data from the courts

Case law is the record of the judgment given in court cases. In essence, it is the 'handing down' of the judgment which becomes part of the body of law. Only a small proportion of judgments in the UK are reported by traditional legal publishers; those reported usually raise a point of legal significance or modify an existing principle of law.³³

Users can choose to access case law through commercial databases, such as Westlaw, or freely available resources such as British and Irish Legal Information Institute (BAILII)³⁴, or directly from the

Supreme Court³⁵ and House of Lords³⁶. At present, case law is not available as open data for England and Wales, so there are inconsistencies with how the foundational data is made available to all.

Data assets

A variety of data inputs related to case judgments include, but are not limited to: parties to the dispute, their details, their legal representatives' details and submissions, the court or tribunal hearing the dispute, the judge(s) or magistrate(s),

³³ www.slideshare.net/acarritt/finding-case-law

³⁴ www.bailii.org

³⁵ www.supremecourt.uk/decided-cases

³⁶ www.publications.parliament.uk/pa/ld/ljjudgmt.htm

the dates and key events in the case, Domestic Acts of Parliament and Statutory Instruments relevant to the case, and European legislation and case law relevant to the case.

Technologies and processes

Some of this data is collected and made available via the web, though not always straight away; however, elements often remain discrete, instead of being linked via unique identifiers. This makes it hard to find accessible data that can inform people as to the complete lifecycle of a case. Despite the volume of data, the question also remains as to whether enough variables are collected to allow the system to be monitored to check that it is working for everyone in society; for example, demographic and diversity data relating to legal counsel, or outcomes, awards and decisions.

BAILII is a non-profit charitable trust that operates and maintains an interactive database of full-text primary legal materials, free to access, online. The list of sponsors who make BAILII's service possible is available on their website, including a mixture of Law Societies, legal publishers, barristers' chambers, academic institutions, and solicitors' firms. BAILII compiles data from free sites, direct and indirect feeds from courts, government departments, and other organisations. BAILII standardises their format(s) to make the data searchable, but this data is not open (licensed and machine readable).

Not all judgements are sent to BAILII. Transcription companies (covered in more detail under Case Study 3) that produce some of the data also send it to legal publishers, who add their own enhancements before publishing it as a paid-for, premium service. This means that access to comprehensive case law data is still geared towards the needs of legally qualified users.

Case law data has made considerable progress in moving from analogue flows to digital; paid for services like Lawtel³⁷ and Westlaw compile much of it in digital format, and provides an easily digestible, quick summary abstract alongside the data. However, data flows in case law still tend to begin as analogue. Although Her Majesty's Courts and Tribunal Service (HMCTS) has renewed its commitment to 'digital by default', its ability to create and manage matters end-to-end in a digital format is currently limited to segments of the judicial system, such as the commercial courts operating out of the Rolls Building in Central London.

Analogue origins dictating digital data flows

This situation can be traced back to the fact that the system of court reporting in the UK is predicated on ways of working that originated with analogue data flows – or paper and ink.

Law reporting remains a crucial part of a transparent system built on the rule of law, which in turn underpins democracy, but its funding and operation reflect print publishing models, with the dissemination of the data eventually tied back to its publication in print. Due to this one-off, print-friendly design, reports tend to focus on such things as the end of the process, giving a transcript of the occurrence, submissions, evidence, and judgement.

If the justice system is being reimagined to be 'digital by default', it seems appropriate to ask – should court reports still be constructed this way? What should the role of court reporters, judges, legal publishers and legal information providers be in a judicial system with digital and data transparency at its core?

Interpretation for different users

To fully unlock the opportunities associated with legal data, new and existing systems interacting with this data will need to cater not only for experts but for non-legal users as well. As well as ensuring that data can be collected, managed and shared easily, it must also be borne in mind that current legal information systems are rarely designed for non-legal users. For example, a digital-by-default justice system must take into account that people looking for information on the court process or their individual case aren't necessarily experts in the legal process.

So, as accessibility to the foundational data becomes digital by default this will open up opportunities for a range of interpretation layers on top. There will continue to be a need for services that combine legal information and layer expert insight on top of the foundational data, but opening up access also creates opportunities for other organisations to provide interpretation for non-expert users. For example, there is a clear need for organisations to provide services and interpretation for those who could not otherwise afford to access legal advice. A more accessible data layer will make it easier for charities or not-for-profit organisations to create their own interpretation services.

To fully unlock the opportunities associated with legal data, new and existing systems interacting with this data will need to cater not only for experts but for non-legal users as well.

Opinion piece: freemium platforms can play a major role in the Open Law Movement

Hugh Logue, Director & Lead Analyst, Outsell, Inc.

Decades of insufficient government funding for public legal information has made it surprisingly difficult for people to easily access the body of case law that governs their daily lives. While legislation is easier to access via government websites and other sources, it is of limited value unless one knows how courts will apply it. The state-run and non-profit public source for case law, such as PACER in the US or BAILII in the UK, have difficult search interfaces, especially for non-lawyers. However, the increased commoditisation of legal data driven by the harvesting of more data sources, is powering a new generation of “freemium” models that provide limited free access to legal research content.

The most ambitious of these is the Caselaw Access Project at the Harvard Law School Library, which is making the largest academic law library in the world available to the public. Its collection has over 42,000 volumes that contain approximately 40 million pages of court decisions, and is the most comprehensive and authoritative database of American law and cases available anywhere except for the Library of Congress.

The project began in 2015 as a partnership that Harvard entered into with Ravel Law, a legal research and analytics start-up. In return for access to Harvard’s law database, Ravel funded the cost of Harvard Law School’s print collections to be digitised by removing the bindings on the case law books and scanning the pages with a high-speed scanner. By the end of 2017, all the processing will be complete, making 334 years of American case law available to the public. Harvard’s materials will be searchable through the Ravel Law platform, which provides visualisations, analytics, and smart-crowd collaborative tools. The advanced judicial and analytics capabilities that Ravel has built on top of the data are fee-based premium features. Harvard can make the data available to the public, with restrictions on bulk access, and make it available to the Harvard community and scholars in bulk for research purposes. Within eight years, Harvard will also release the entire database for bulk use by anyone, including launching an API for other commercial information providers to use.

In June 2017, legal research provider LexisNexis announced the acquisition of Ravel Law. In public statements at the time of the announcement, LexisNexis and Harvard committed to continue maintaining the free legal research platform. LexisNexis will use Ravel Law’s technology to improve its legal analytics capabilities.

Ravel is not the only start-up that used the freemium model to increase access to law by the public. Others include:

- **Casetext** provides a free legal research platform that uses crowdsourcing to identify important cases and offer legal analysis. While simple search and retrieval of cases is free, the use of advanced feature such as CARA, an AI tool for analyzing legal briefs, requires a subscription.
- **Co/Counsel** is a crowdsourcing legal research platform that uses graphic interfaces to serve both lawyers and laypeople. It visually maps law by organizing cases, statutes, and articles into logical interactive flowcharts with similar legal principles grouped into factors and sub-factors. It sells its solutions to the academic market, but makes its legal maps free to the public after the semester is over.
- **Justis Publishing** recently launched a legal public resource in the UK through a freemium research platform, JustisOne. The offering covers case law from 2001 and includes basic search tools, while the premium version covers case law from 1855 and legislation from 1235. The strategy provides Justis with greater online discoverability by search engines, which in turn drives sales.
- **vLex** offers a free Chrome browser extension in Spain that enables lawyers to use vLex’s law citator on any webpage on the internet. The tool helps vLex to identify new prospective customers.

In theory, this new wave of free legal content providers threatens commercial legal research platforms, such as Bloomberg Law, LexisNexis or Westlaw. This will particularly be the case once Ravel’s commercial license with Harvard ends after eight years, and the data will become available for exploitation by other organisations and businesses. However, case law on its own is difficult to grasp without some context and direction from experts. As more people gain access to free case law, there will be more demand for explanatory content and analytics tools, which creates new opportunities for legal solution providers. The value of legal research platforms is no longer in the data, but in the tools that analyse and decipher that data.

Opinion piece: Nick West, Chief Strategy Officer, Mischon de Reya LLP

It's extraordinary that we're even having a debate in 2018 about whether and how court case data should be open. The Rule of Law requires transparency of law and since the law in England & Wales is (partially) made by judges sitting in court interpreting statute, it follows that court decisions and other information about court cases should be freely available to the public.

Yet, the reality is very different. One can freely access a couple of decades of Supreme Court and Employment Appeal Tribunal decisions, plus the hodgepodge of cases in BAILII. There currently exists no central free-to-air database of court activity (cause lists and so on)—not even current, let alone historic—and no central database of court decisions. Even acknowledging that the sheer volume of cases in the Magistrates and County Courts and lower tribunals presents a challenge, there's not even an easy way to access most of the current and historic data from the higher courts and tribunals.

By contrast, in many countries around the world the situation is different. The public has easy access to court data and decisions. These jurisdictions have grappled with, and overcome, issues about central IT infrastructure, timing of publication, rules preventing sensitive information being published and several other apparent hurdles. The result? Greater access to justice, as information asymmetry is reduced; and greater innovation through connected datasets enabling predictive analytics.

People have been talking about this issue—open data and access to justice—for more than 10 years. The legal profession and legal publishers won't provide the solution—the incentives are all wrong. It must come from Government. With Brexit looming, there is significant work being done to trumpet our world-leading legal services sector and ensure it remains at the head of the pack. Isn't it time we had world-class data infrastructure to support it?

Case study 2: court listings

Court listings information details the cases which are scheduled to be heard by a particular court on a particular day, or other period of time. A typical court listings record might include the court number, the case number, the defendant's name and current status of the case. Court listings are an important source of data, underpinning a transparent legal system.

Court listings represent a data area where efficiencies and improvements in data sharing must be balanced with appropriate data governance principles regarding personal information.

Data assets

The facility for the public dissemination of listings data is not consistent across UK courts and tribunals, and is not open data:

- Magistrates' courts produce listings every Friday with data such as party's name, date of birth, address, and alleged charge.

- Crown Courts update listings information on a daily basis, accessible through a government website.³⁸
- The Royal Courts of Justice and the Rolls Building list their hearings on a different government webpage.³⁹
- A range of outlets provide County Courts listings in some format or other, including County-Courts.co.uk⁴⁰ and CourtServe⁴¹.

What is clear is that the level of detail, and frequency of updates, of court listings in England and Wales is neither complete nor consistent, although much progress has been made in listing online in recent years. Listings information is available to journalists either through a local presence or through having the resources to maintain a presence in the relevant courts across the country. Magistrates' court listings can be requested by local newspapers in email format,

38. xhibit.justice.gov.uk/court_lists.htm

39. www.justice.gov.uk/courts/court-lists

40. www.county-courts.co.uk

41. www.courtserve.net

or in hard copy by collection, though the cost and the procedure vary.⁴² This does not represent effective data infrastructure.

Balancing accessibility, transparency and privacy

To make this data findable, accessible, interoperable and reusable, there would be a broad benefit from accessing court listings in a consistent way with expectations around timeliness and formatting. However, according to the principles of the data spectrum, it would not be appropriate to make all information available as open data to everyone. The procedure of providing listings to journalists from Magistrates courts is a way to deal with confidentiality. Some sensitive data, particularly address and date of birth is not appropriate to be open, and data systems must appropriately protect individual rights.

Furthermore, there is the fact that particularly in criminal cases, a listing represents an allegation, which may turn out to be unfounded. Similarly, in civil litigation, the given dispute or claim may well be resolved in a way that absolves one party

completely. However, if the data pertaining to the case's outcome is harder to access than the listing data, it can give an unbalanced picture of the involvement of an individual or a company in a legal action. The accusation or claim may be available to all, but the rebuttal of it may remain relatively unreported.

As one considers the skills or services necessary to understand, this is where linked records and a complete picture of a court case can keep information in context. This could also support an individual in their 'right to be forgotten'⁴⁴.

Newer systems, like the CE-File solution developed at the Business and Property Courts/Rolls Building⁴⁵, capture the full lifecycle of a case, through listing and evidentiary stages right through to the verdict and judgement, all in digital format. However this is, for the moment, an exception in the legal system of England and Wales, rather than the rule. In redesigning court listing systems to facilitate open data and improved access to justice, we must bear in mind the data points that need to be recorded at all stages in order to enable users eventually to explore and search amongst ongoing cases.

Some sensitive data, particularly address and date of birth is not appropriate to be open and data systems must appropriately protect individual rights.

Case study 3: case transcription data

Data assets

Audio recordings are made of many court proceedings and can be requested by way of an application to the presiding judge.⁴³ Transcriptions of court proceedings are handled by a number of contractors to the Ministry of Justice, and form the basis of the outputs made available by BAILII and other data providers, along with data gathered before court proceedings—such as legal advice, pre-trial documents, or details of the police procedures leading to an arrest or charge.

Access varies from court to court, partly due to the fact that transcription must be driven by demand. Making original recordings available more widely

would be inappropriate because of those occasions when sensitive information is shared in court that should not be openly available. Litigants may not even be aware of the existence of a recording or transcription of their proceedings, let alone any potential right they have to access it. Participants in court proceedings may request transcriptions via a specific form⁴⁶, listing the authorised members of the transcription panel, who will subsequently invoice the party making the request.⁴⁷ Costs can be substantial, with some appeals proceedings quoting £105 per hour for transcription services.⁴⁸ But despite the hurdles and costs associated with transcription, it generates significant and valuable data for users of the court system.

42. www.newsmediauk.org/write/MediaUploads/PDF%20Docs/Protocol_for_Sharing_Court_Documents.pdf

43. ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/the-right-to-erasure

44. www.gov.uk/guidance/ce-file-system-information-and-support-advice

45. www.judiciary.gov.uk/publications/practice-direction-access-to-audio-recordings-of-proceedings

46. formfinder.hmctsformfinder.justice.gov.uk/ex107-eng.pdf

47. formfinder.hmctsformfinder.justice.gov.uk/ex107-info-eng-20160121.pdf

48. legal.ubiquis.co.uk/civil-transcripts

Opinion piece: deploying technology to speak truth to justice

Rachel Mpashi-Marx, Director, Just:

In the movies, miscarriages of justice are often resolved through exciting coincidences, acts of bravery, or the work of a maverick lawyer. The reality more often is that success lies in the paperwork - hidden in the piles and boxes of documentation may be the key to unlocking the case. Frequently, one of the most important documents is the transcript of the original court hearing.

In America, one of the first things the lawyer does with a miscarriage of justice case is get hold of the transcript to find out what went on in court, and where things went wrong. Often, it is the only decent source of information available.

In the UK, working with the Centre for Criminal Appeals, getting hold of court transcripts proved almost impossible. Audio recordings are held by a confusing maze of private companies, and the high costs and inconvenient timings are major barriers. One of our clients was quoted £20,000 for the transcript of the hearing that convicted him of an attempted murder he had no part in.

The reasons for this are several. The Ministry of Justice (MoJ) has a series of contracts with private transcription providers, many of which are part of for-profit multinational companies. The intellectual property (IP) for the transcripts lies not with anyone involved in the case, or the State, but with these private transcription providers. Transcripts are only produced on request from a paying customer, and only available for the very highest courts. Wealthy commercial clients have found a way around this problem - they pay, often up to £800 per day, to have a typist in the court. For most people, of course, this is just not an option.

The service that the MoJ commissions has hardly changed in years. Transcripts are routinely made from audio files burned onto CDs and couriered around the country.

The process for requesting a transcript is complicated and opaque. The unfriendly EX107 form needs to be filled in, printed off, signed and then scanned or posted off. It is left up to the person requesting the transcript to select a provider. If your request is approved, the transcription provider gives you a quote for the cost of producing the transcript, and form EX107 states, 'you may wish to negotiate a price within the maximum with the transcriber.' For seasoned lawyers used to requesting

transcripts, this may be straightforward. For a stressed out litigant-in-person who has never done this before this does not seem reasonable.

There are no specific guidelines about how long these audio files should be retained, beyond standard data protection rules. This means that tapes and CDs are being destroyed after six to seven years, whether or not that makes sense in terms of the case itself, and without the people concerned being informed. This means that if an inmate might benefit from, for example, an advance in science that enables a fresh appeal many years down the line, there may be no record of what happened originally to work from.

The service that the MoJ commissions has hardly changed in years. Transcripts are routinely made from audio files burned onto CDs and couriered around the country. These are then listened to and hand typed by teams of typists. Rumour has it, some audio recordings are still stored on cassette tapes.

What if transcripts were automatically generated using speech-to-text transcription technology, and the documents centrally held? What if the IP were back in the hands of the MoJ, and the information managed on open data principles? What if the resource burden could be reduced to such an extent that transcripts were produced as part of business as usual, even for the lower courts and tribunals? What if we could analyse the content of the transcripts to gain unprecedented insight into what is happening across our justice system? If all of these were true, we would have a means of understanding what is actually going on in our courts that can currently only be dreamed of.

As more digital resolution options become available, what does a transparent justice system look like? With the advent of online courts, what is the digital equivalent of the public gallery? Transcripts have an important role to play in keeping the justice system in the light as new possibilities take shape. Now is the time to plan for such an outcome. It will be much harder to implement retrospectively. The legal sector is coming relatively late to this party, so much of the technology is already tried and tested in other industries. With so many innovative technologies available, as well as growing pressures on the system from funding cuts and higher expectations about service provision, there has not been a more appropriate or exciting time to re-think how we document our legal system.

Case study 4: legislation

Legislation, or ‘statutory law’ defines both the act of making and forming law, and the laws formed themselves by the legislative process.

UK domestic legislation is openly accessible through legislation.gov.uk, operated by The National Archives. A number of similar resources exist internationally: EUR-Lex⁴⁹ provides access to European Union law, and Congress.gov⁵⁰ provides access to U.S federal legislative information. Accessibility is probably more advanced for legislation when compared with other parts of legal data infrastructure. When it comes to interpretation for legal professionals, commercial resources such as Westlaw add expert guidance and interpretation on top of the foundational data.

The National Archives publish legislation as it is enacted by Parliament, and to a certain extent the changes it makes to existing legislation and the result of these changes.⁵¹ It does this online through legislation.gov.uk—a unique public service built on collaboration between the public sector and private sector, and open data.⁵² Access to the government’s legislation database was opened up by creating an API first then building the legislation.gov.uk

website on top of it. Developing the API enabled the team to tackle a backlog of amendments and take an entirely new approach to providing more up-to-date legislation—inspired by the open source movement—their Expert Participation Programme.⁵³

By changing the process to update legislation and by providing tools to the community, participants from private and public sector joined the programme, willing to contribute to the process to reap the rewards of high quality and up-to-date open data. Recruited into the programme were 72 participants from 10 different organisations and 51 percent of the legislation updates were done by people outside of the National Archives.⁵⁴

In addition to the volume of update work completed, other impacts include breaking down of silos across government, and the quality of the data itself—additional facts, such as the geographic extent of the amendments are now recorded within the data.⁵⁵ The programme and processes adopted by the legislation.gov.uk team provide a clear methodology for anyone that owns and maintains a large dataset and is looking to transform their existing processes.

When it comes to interpretation for legal professionals, commercial resources such as Westlaw add expert guidance and interpretation on top of the foundational data.

49. eur-lex.europa.eu/homepage.html?locale=en

50. congress.gov

51. theodi.org/case-studies/case-study-legislationgovuk

52. *Ibid*

53. gds.blog.gov.uk/2012/03/30/putting-apis-first-legislation-gov-uk

54. theodi.org/case-studies/case-study-legislationgovuk

55. *Ibid*

The impact of Brexit on legislation

John Sheridan, National Archives, and Amanda Smith, Senior Consultant, Open Data Institute

Exiting the EU poses complex challenges to the legal sector. The UK needs to both retain and modify the large body of European law it has followed since 1972.

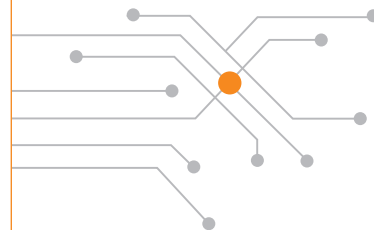
To cope with this challenge the UK Government will be able to use open data from the EUR-lex website and the CELLAR database (a content and metadata repository) both developed by the Publications Office of the European Union, as well as its own data from legislation.gov.uk and value-added services from commercial publishers.

Having a strong legal data infrastructure is crucial in a situation like this. The government needs to make sure there is certainty about what law applies when the UK leaves the EU. Government departments need to know and understand what laws need to be retained and changed as part of Brexit. To put the challenge into context, the Department for Environment, Food and Rural Affairs (Defra) has approximately

80 per cent of its work framed by EU legislation, and estimates a quarter (1,200) of EU laws relate to their sectors and areas of responsibility.

The European Union (Withdrawal) Bill⁵⁶ creates a new duty for the publishing of all the relevant retained EU legislation and a power to publish anything else useful related to it. The National Archives, who are responsible for legislation.gov.uk, will be publishing all this information using open data from the EU as a source.

Legal data infrastructure at its most basic use helps the functioning of Government: it helps departments to understand what regulations they are responsible for, to make the changes necessary to the law for the country to function once the UK leaves the EU, and for all the information to be in the public domain so people and organisations know where they stand.



56. www.publications.parliament.uk/pa/bills/cbill/2017-2019/0005/cbill_2017-20190005_en_4.htm#sch5

Case study 5: bringing other data assets to the legal profession

Data benefits from network effects: it creates more value as more people use, contribute to and maintain it.⁵⁷ The earlier examples of data infrastructure in this paper were those that are unique to the legal sector, but in this final section we look at other examples that form part of the data of record that underpins legal work, but are not unique to legal use; such as from the Land Registry and Companies House.

The data from these sources is often, by definition, open and accessible to anyone, but within the legal sector there is a particular opportunity for citizens, businesses and their legal advisers to benefit from accessing this data, and identifying new valuable sources of data from other sectors. The potential is to increase efficiency and reduce cost in the legal system by drawing together the whole data spectrum when addressing legal issues.

Land Registry

For England and Wales, the Land Registry maintains the records of land and property ownership, the land charges registers, the bankruptcy register and the agricultural credits register within a remit set by legislation and policy. Approximately 24 million property titles (the evidence of property and land ownership) across England and Wales are included in these records. For the rest of the UK there are the Registers of Scotland⁵⁸ and the Land & Property Services⁵⁹ (housed within the Department of Finance) for Northern Ireland.

Datasets published⁶⁰ by Land Registry include price paid data⁶¹, transaction data⁶² and the UK House Price Index (UK HPI)⁶³. Property information is available at a charge (from £3 per search) which clarifies who owns a property, what they paid for it and whether there's a mortgage on it⁶⁴. Whilst its

HOW IS LAND REGISTRY DATA BEING USED?

Shared Assets⁶⁵ undertake research and policy work and are advocates for good land use and change in land management approaches. Their online platform, Land Explorer⁶⁶ uses open data from Land Registry, as well as the Environment Agency and Natural England. It provides users with information in a clear, interactive way to help them find land, understand the land in their local area, inform planning applications and protect environmental and community assets.⁶⁷ Land Insight⁶⁸, DealX⁶⁹ and The Land App⁷⁰ are just some of the startups innovating in the PropTech sector, offering services to discover and visualise land, understand and assess development opportunities, and plan projects.

'Who owns England?'⁷¹ use the data to shine a light on what land is owned by the Duchy of Cornwall⁷², which 50 companies own over half a million acres of England and Wales⁷³, or how to solve land banking⁷⁴ (a practice where land is hoarded without being developed in the hope that it can be sold on for a premium price).

In 2016, Thomson Reuters Labs partnered with Transparency International to look at corrupt wealth in London property. This drew on Land Registry data as well as the Panama Papers, data sources from Thomson Reuters and OpenCorporates.⁷⁵

57. theodi.org/guides/principles-for-strengthening-our-data-infrastructure

58. www.ros.gov.uk/

59. www.finance-ni.gov.uk/land-property-services-lps

60. www.gov.uk/topic/land-registration/data

61. www.gov.uk/government/collections/price-paid-data

62. www.gov.uk/guidance/land-registry-transaction-data

63. www.gov.uk/government/collections/uk-house-price-index-reports#about-the-uk-hpi

64. www.gov.uk/search-property-information-land-registry

65. www.sharedassets.org.uk/what-we-think

66. landexplorer.uk

67. www.sharedassets.org.uk/landexplorer/uses-for-land-explorer

68. www.landinsight.io/about

69. dealxlab.com

70. www.thelandapp.com

71. whoownsengland.org

72. whoownsengland.org/2017/03/15/what-land-does-the-duchy-of-cornwall-own

73. whoownsengland.org/2016/12/12/the-50-companies-that-own-over-a-million-acres-of-england-wales

74. whoownsengland.org/2017/02/07/how-to-solve-land-banking-reveal-where-its-happening

75. www.transparency.org.uk/publications/london-property-tr-ti-uk

core services underpin the property market and new innovations in the 'PropTech' industry, the data that the Land Registry holds and publishes is used in many other sectors.

In 2016, the Government launched a consultation⁷⁶ on privatising the Land Registry as part of the Neighbourhood Planning and Infrastructure Bill. The primary motivation behind this proposal was to allow government to identify up to £5 billion of additional corporate and financial sales assets (by 2020) to pay down debt⁷⁷. Following criticism from the Competition and Markets Authority (CMA)⁷⁸, the UK Open Government Network⁷⁹ and others (including the ODI⁸⁰ and a public petition which attracted over 320,000 signatures⁸¹) that this sale would compromise the services provided by Land Registry, the Government has cancelled its plans to privatise the Land Registry.

In the 2017 white paper, 'Fixing our broken housing market'⁸² Land Registry have committed to becoming the world's leading land registry for speed, simplicity and an open approach to data. The latter is crucial as it guarantees not only access to the data, but also ease of reuse. Land Registry will make its commercial and corporate ownership data set and the overseas ownership data set available free of charge. These data sets contain 3.5 million titles to land held under all

ownership categories (with the exception of private individuals, charities and trustees). Both datasets were published in November 2017, providing valuable insight to lawyers and businesses in the commercial property sector.

Companies House

There are more than 3.5 million limited companies registered in the UK and 500,000 new companies incorporated each year. Companies House incorporate and dissolve limited companies, examine and store all company information and make this information available to the public in the official government register of UK companies⁸³, which contains 170 million company records⁸⁴.

All of the public data held on companies is made available, free of charge, through the Companies House Service⁸⁵, with an API⁸⁶ to encourage reuse. Information available through the service includes basic company information, the nature of the business, company status, dates of last and next accounts/confirmation statements, mortgage charge data and insolvency information.⁸⁷ The company ID created by Companies House can be used as an anchor to link data across your different sources. In essence, Companies House enables people to discover many interesting things about relationships, to understand and create connections.

HOW IS COMPANIES HOUSE DATA BEING USED?

Providing information like this supports the economy; it enables people to compare the companies they are going to do business with and use it to make better decisions. It drives growth, and uncovers fraud. The millions of public records published by Companies House are used by police investigators, the National Crime Agency, Serious Fraud Office, lawyers, journalists and bank compliance teams in searching for information about dissolved

companies and their directors to tackle money laundering and white collar crime.⁸⁸

OpenCorporates—the largest open database of companies around the world – use Companies House data to make information on companies more usable and more widely available for the public benefit, particularly to tackle the use of companies for criminal or anti-social purposes, for example corruption, money laundering and organised crime.⁸⁹

76. www.gov.uk/government/consultations/land-registry-moving-operations-to-the-private-sector

77. www.gov.uk/government/uploads/system/uploads/attachment_data/file/510987/BIS-16-165-consultation-on-moving-land-registry-operations-to-the-private-sector.pdf

78. www.gov.uk/government/publications/cma-response-to-bis-consultation-on-the-future-of-the-land-registry

79. www.opengovernment.org.uk/2016/05/24/land-registry-privatisation-threatens-open-government-support-the-ogns-consultation-submission

80. theodi.org/blog/consultation-response-on-moving-land-registry-operations-to-the-private-sector

81. you.38degrees.org.uk/petitions/stop-privatisation-of-the-land-registry

82. www.gov.uk/government/uploads/system/uploads/attachment_data/file/590463/Fixing_our_broken_housing_market_-_accessible_version.pdf

83. www.gov.uk/government/organisations/companies-house/about

84. companieshouse.blog.gov.uk/2015/06/22/free-access-to-over-170-million-company-records

85. beta.companieshouse.gov.uk

86. developer.companieshouse.gov.uk/api/docs

87. www.gov.uk/government/organisations/companies-house/about/about-our-services#chs-info

88. www.theguardian.com/business/2016/aug/02/companies-registrar-under-pressure-to-wipe-information-from-database

89. opencorporates.com/info/about

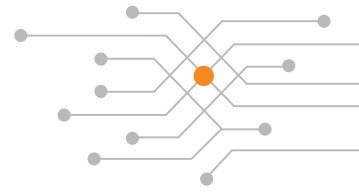
Conclusion

The legal system is based on large volumes of written information, and we have proposed the view to think of this as data. Following this view, in turn, one can see how applying the principles of good data infrastructure can aid lawyers, citizens, businesses and governments.

In this paper, examples were provided for the existing ways in which public and private sector organisations, communities and citizens of most countries, have a shared interest in strengthening their legal data infrastructure. The Brexit process in the UK highlights that access to the right

legal data will become even more important, but this need already exists, is global, and extends far beyond Brexit. The legal sector contributes significantly to the global economy. Doing business and living in a globally connected, changing world means that it is vital that we strengthen our legal data infrastructure.

Whilst this paper did not set out a blueprint for that process, our aim is to further the conversation to help achieve an appropriate level of advancement for the global legal data infrastructure.



Acknowledgements

We would like to thank the following individuals and organisations for their time and contributions to this paper:

AUTHORS

Open Data Institute

Amanda Smith, Senior Consultant

Dr. Jeni Tennison, CEO

Leigh Dodds,
Data Infrastructure Programme Lead

Thomson Reuters

Dr. Tharindi Hapuarachchi,
Technical Partnerships Manager,
*Thomson Reuters Labs*TM

Dr. Andrew Fletcher, Director,
*Thomson Reuters Labs*TM

Desmond Brady,
Director, Public Sector, Academic and Bar

OPINION PIECE CONTRIBUTORS

Marie Bernard, Europe Director of Innovation,
Dentons

David Curle, Director, Market Intelligence,
Thomson Reuters

Hugh Logue, Director & Lead Analyst,
Outsell, Inc.

Nick West, Chief Strategy Officer,
Mischon de Reya LLP

Rachel Mpashi-Marx, Director, *Just:*

John Sheridan, Digital Director,
The National Archives

Contact Us

Thomson Reuters

If you'd like to receive more reports and exclusive content about legal industry trends from Thomson Reuters visit us at: legalsolutions.thomsonreuters.co.uk or: blogs.thomsonreuters.com/legal-uk

Follow us on Twitter: [@TRLegalUKI](https://twitter.com/TRLegalUKI)

Follow us on LinkedIn, search:

Thomson Reuters – Legal UK & Ireland

Email: legalsolutions.uki@tr.com

