Open Addresses: the story so far

Executive summary

Open Addresses was initiated in July 2014 and launched in January 2015 in response to the sale of the Postcode Address File (PAF) as part of Royal Mail's privatisation. It was set up to explore how an open address database for the UK could be created using a collaborative model. It was supported by the Open Data User Group (ODUG) and funded by the Cabinet Office, through the Release of Data Fund.

In this report we aim to help others to learn from Open Addresses **mistakes and successes** to date. We also highlight the considerable **challenges** that those looking to innovate with address data are likely to face, including those presented by the current **uncertainty around intellectual property rights** in data.

While the project is still in its early stages as a business, it has already revealed some important lessons:

- The legal situation around address data is complex and unresolved, making innovation in the sector extremely difficult
- Existing address services in the UK can be improved using open and collaborative approaches, and these services are often more important than complete address databases
- The authority of existing address providers may be more important to the existing market than openness, transparency or agility, but there are new markets for which it is less important
- The service-based framework for a sustainable business model seems tenable, but it is too early in the project to substantively test the market
- There are no substantial technology issues; this is not a blocker to development

The project also raised challenging questions at four levels:

- **Legal**: what are the lawful limits on the reuse of data published under open licences?
- **Governance**: what characteristics do we expect in entities that maintain nationally important information infrastructure, to build trust that the data they hold will continue to be available and of high quality?
- **Technical**: what kinds of system infrastructure needs to be built to support crowd sourcing of high-quality data?
- **Business model**: how can organisations that provide free access to open data make money in order to be sustainable?

National Information Infrastructure should be maintained by organisations that (1) can be trusted to exist in the long-term, (2) have a reason for continuing to make data available to others, and (3) can adapt to changing user needs and expectations.

Businesses, organisations, government and members of society need to be able to rely on access to sustainable, high quality data in order to build new business models, plan better services and products and gain insights into how we live.

Open Addresses Limited, the company created to see through the Open Addresses project, does not meet all these criteria, but illustrates an alternative governance approach to those demonstrated by Royal Mail and GeoPlace: one fundamentally based on openness, transparency and collaboration.

Business models for companies that provide open data are different from those predicated on selling licences to use data. But the need to build services that add value to data can improve the experience for customers and end-users. Open Addresses identified a **freemium model for the provision of address-based services** such as parsing, auto-completion and matching that (1) are not reliant on an authoritative set of address data (2) can gradually improve the coverage and quality of address data and (3) improve user experience in address entry and usage.

In the UK, data can be protected under both copyright and database rights. **Open data** promises simplicity: it can be **accessed, used and shared by anyone**. However, open licences cannot license intellectual property rights held by third parties and in practice it can be difficult for data managers to identify whether third-party intellectual property exists in their data.

For example, it is not clear from existing statutory or civil law whether acts such as validation (eg checking if an address exists in an authoritative list from a third party) **embed intellectual property into existing data**. Open Addresses encountered many instances of data published under an open licence that were potentially contaminated with third-party intellectual property.

To **minimise risks of copyright and database right infringement**, Open Addresses avoided these datasets and put processes in place to remove such data if infringements came to light. This approach limited the ability of Open Addresses to incorporate public sector data - collecting 1.7m addresses out of about 30m available within open data - and introduced technical complexity by requiring detailed provenance records. Even with these precautions there is a risk of legal action being taken by Royal Mail; a risk that insurers judge as too high for them to cover.

There are several **models for crowd-sourcing data**. OpenStreetMap offers a free-for-all where deliberate or accidental errors are corrected by the crowd. Legislation.gov.uk requires its contributors to be trained and their work reviewed. Open Addresses reused a moderated sandboxing approach for third-party scripts originally developed by OpenCorporates; it provided **small APIs** for activities such as parsing and validation that help to **improve the quality of data**; and it built services that help reusers assess the confidence they should have in contributed address data.

Open Addresses is an important project. As well as aiming to provide address data as open data, it thoughtfully explores new territory.

Acknowledgements

Table of contents

Introduction

> *"There is no doubt that addressing – the network of road names and house numbers – constitutes a key element of functioning societies. While a single address in itself does not constitute a public good, the national address infrastructure, of which it forms a part, is an essential public good, and through interoperability with international systems the totality of addressing networks can be determined as global public good.*
>
> *Address infrastructure provides access to the rights and duties of citizens from the local to the international level, as well as providing businesses with access to markets. All echelons of society should thus have equal access to address infrastructure in order to capture the social and economic benefits at the local, national and international levels."*

<div align="right">

Addressing the World - An Address for Everyone White Paper, Universal Postal Union
Berne, Switzerland (2012)

</div>

Benefits of open address data

Address data has **significant economic value**. It underpins products and services across every sector of the UK economy. A 2012 report from the PAF Advisory Board estimated its value to the UK economy to be between £992m - £1.32bn each year.

In 2002, the Danish Government made its national address file available for free. Since then it has calculated its direct financial benefits to cost ratio at 30 to 1. By 2010, the year in which their report was published, they estimated the direct benefits of free address data to the Danish population to be around EUR 14m, with costs of EUR 0.2m. Scaling to the UK population, this would be an annual benefit of around £110m.

In their 2014 report for the Department of Business, Innovation & Skills, Katalysis recognised the need for Open National Address dataset:

> *"UK society relies heavily on address data and current products have helped greatly to create benefit. The review has determined that Open usage would result in substantial and valuable growth among new user types and with even greater community benefit. The recommendation is that a basic address product should be free to all users at the point of use."*

The benefits of an open address database for the UK were reinforced by the Open Data User Group (ODUG), who said,

> *"The benefits [...] include improvements in the transparency and delivery of public services, benefits to business innovation and growth, immediate cost savings as a result of streamlining address data aggregation and delivery mechanisms, and improved address data quality as a result of a single central data repository."*

Issues associated with construction of the address register for the 2011 UK Census indicate the need for an open addresses database. Construction was laborious and cost

£3m. The register has [never been released](#) nor kept up-to-date, because of the restrictions in place on the use of Royal Mail and Ordnance Survey intellectual property. An open address database would enable activity to be focused on an **address database that would be public property, available to all** and maintained collaboratively.

The sale of the Postcode Address File

> **"The sale of the PAF with the Royal Mail was a mistake.** *Public access to public sector data must never be sold or given away again. This type of information, like census information and many other data sets, is very expensive to collect and collate into useable form, but it also has huge potential value to the economy and society as a whole if it is kept as an open, public good."*
>
> UK Government Public Administration Select Committee (PASC) report on Statistics and Open Data, Chair Bernard Jenkin (2014)

The Postcode Address File (PAF) is the UK's authoritative database of postal addresses (or delivery points) in the UK. It contains 1.8 million postcodes and over 28 million addresses. Each day, millions of UK citizens interact with services incorporating PAF data while doing things like shopping online, registering for services and making appointments.

In 2013, PAF was included by the UK Government in its sale of Royal Mail. At the time, the Open Data Institute, the Open Data User Group, and others argued against the inclusion of PAF in the sale, and encouraged the release of PAF under an open licence.

Before being privatised, Royal Mail was known for its litigious approach to services that might encroach on PAF. In 2009, Royal Mail lawyers [issued a cease and desist letter](#) to Ernest Marples Postcodes Ltd, a UK postcode lookup service that provided web providers with an API to help people search for information specific to their area. Tom Watson, Labour politician and former government minister for digital engagement, described Royal Mail's actions at the time as "heavy handed", [saying](#):

> *I take the position that the postcode file and the data set of physical coordinates that go with it are a national asset that should be freely available to any UK citizen.*

In testimony before the Public Administration Select Committee (PAF) in 2013, Sir Nigel Shadbolt [described](#) the value of PAF as:

> *"a common good [...] Almost every conceivable new advance in delivery of services uses digital capability; everything happens somewhere, everything gets delivered somewhere, whether it is blue light services or commercial innovation."*

ODUG [commented that](#) to allow Royal Mail to take PAF into private ownership as a commercial dataset, and for Ordnance Survey to participate in the creation of GeoPlace

LLP as a trading Value Added Reseller of PAF, would appear to "fly in the face of any Government commitment to open data."

Since the privatisation of PAF, ODUG has suggested that PAF could still be released under an open licence. In 2011-2012, licensing revenues from PAF only accounted for [approximately 0.3%](#) of of Royal Mail's revenue. **Making PAF available under an open licence would not significantly impact on Royal Mail's revenue projections**.

Royal Mail could feasibly benefit in reduced delivery costs from an increased quality of addressing that would result from greater adoption of their standard address forms, and more accurate use of postcodes. It could also potentially **reduce its sales and legal expenses by adopting an open model**, and reduce maintenance costs by adopting an **open, collaborative maintenance model**.

However, changing technologies to reduce costs and provide for alternative maintenance approaches would itself require investment. And there is a cultural division with Royal Mail whereby the Address Management Unit, which maintains PAF, is kept separate from other business units, making the cross-subsidisation that would be required less feasible.

What is Open Addresses?

The community of organisations and individuals who are interested in address data agree that a single canonical high-quality open national address database would bring wide benefits to the UK.

Open Addresses aimed to investigate how viable it would be to **create a free and open database of UK addresses** through collaboration, in order to test the two hypotheses:

1. it is possible to build and maintain a sustainable open address database using collaboration, cross-subsidy and volunteer effort, in a similar way to OpenStreetMap
2. the maintenance of open address data can only be effectively funded through taxation, with the implication that current governance models for address data in the UK must change to achieve the full benefit from address data

The first hypothesis is the basis for the Open Addresses project. Failure would imply that the second hypothesis is correct: that the only way to secure an open address database for the UK would be for the government to subsidise its maintenance, probably through an existing organisation such as GeoPlace.

The precise cost of maintaining address data, even using current governance structures, is debatable. Simply subsidising the practices of existing organisations would presumably require government to not only fund the maintenance of the data itself (in addition to the payments that it already makes through the Public Sector Mapping Agreement (PSMA),

including the time already committed by local authorities) but also recompense Royal Mail and Ordnance Survey for their loss of earnings.

Even if Open Addresses failed to attain sustainability, it was hoped that the technologies and processes for collaboration that it developed could be reused by existing organisations to reduce their costs, and therefore reduce the overall cost to government – whatever the eventual scenario.

Funding Open Addresses

The Open Data Institute applied to the Cabinet Office for funds to support the development of Open Addresses through the Release of Data fund in May 2014. The Open Data User Group (ODUG) identified open address data as a priority for the wider open data community, and supported the ODI's application.

The proposal estimated the total cost of bringing Open Addresses to beta release over nine months as £343,000 + VAT (£411,600). The ODI was offered £28,800 to support the Discovery Phase of Open Addresses in July 2014. Open Addresses Ltd was offered £382,800 to support the Alpha and Beta phases in December 2014. This money had to be committed by end of March 2015; spent according to the original proposal, unless Cabinet Office agreed otherwise; and could only support work in England and Wales.

Discovery phase

The discovery phase of Open Addresses, which ran during July and August 2014, included:

1. Researching a suitable **governance model** for Open Addresses
2. Developing an initial **business model and costings** for the ongoing maintenance of Open Addresses, including exploration of **potential products and services** that could be offered by Open Addresses as part of a pricing model
3. Assessing **technical architecture and data models** to support Open Addresses
4. Developing a **communications** plan and **community engagement** around Open Addresses, including delivering the Open Addresses Symposium on 8 August 2014
5. Seeking **legal advice** regarding potential IP and other risks associated with Open Addresses
6. Creating an **initial delivery plan** for the project

Alpha phase

The alpha phase of Open Addresses ran from December 2014 to the launch of the alpha website on 14 January 2015. This included:

1. Incorporating **Open Addresses Ltd** and appointing Non-Executive Directors

2. Creating the initial **technical infrastructure** for Open Addresses, including the website and basic API for data access
3. **Loading 1 million addresses**, sourced from Companies House data
4. Initiating the **communications and community engagement** plan

The beta phase of Open Addresses ran from January 2015. The activity funded by the Cabinet Office grant included:

1. Further **developing the business model and planning** for Open Addresses Ltd, including identifying the services that Open Addresses should offer through market research
2. Continuing work to build out the **technical infrastructure** for Open Addresses, including APIs for parsing, validating, and assigning confidence to addresses
3. Expanding the number of addresses by **improving parsing performance**, to 1.7 million addresses, and building algorithms to infer more
4. Improving the **user experience** of the Open Addresses website
5. Conducting **user testing** to identify the interfaces that best support end-users with different levels of digital experience to enter addresses
6. **Community outreach and engagement** through a photo competition, sponsorship of Open Data Camp 2015, and various speaking engagements
7. Writing this **lessons learned** report

Almost £10,000 was also provided in April 2015 by the [British Computing Society (BCS)](), The Chartered Institute for IT's Location Information Specialist Group's (LISG) [Digital National Framework]() fund, specifically to pursue the availability of URLs for addresses and the use of Unique Property Reference Numbers (UPRNs) within open address data.

It is worth noting that there were considerable delays in the process that had a knock-on impact on the delivery of Open Addresses. The original bid contained a nine-month development plan, with one month for discovery and eight for alpha and beta development. The discovery phase was actually delivered within a month, but delays in the sign-off of the remaining grant meant that alpha and beta development was compressed into five months. Creating a new entity also entailed a high administrative burden: for example, it took three months to open the bank account for Open Addresses.

Timeline: The road to Open Addresses

- [December 2010]()
  GeoPlace LLP formed by Ordnance Survey and Local Government Association. It's role: to create and maintain the National Address Gazetteer and the National Street Gazetteer for England and Wales, providing definitive sources of publicly-owned spatial address and street data

- April 2011
  Public Sector Mapping Agreement initiated, providing UK public sector bodies access to Ordnance Survey products
- Autumn 2011
  National Address Gazetteer created, bringing together Local Land and Property Gazetteers, Address Layer 2 and Royal Mail PAF data
- January 2013
  Public Sector Transparency Board discussed Open Data User Group business case for open address data
- March 2013
  Sir Tim Berners-Lee and Sir Nigel Shadbolt write to the prime minister to argue for opening the Postcode Address File (PAF)
- October 2013
  Postcode Address File (PAF) sold as part of the privatisation of Royal Mail
- November 2013
  ODI submits original proposal for funding Open Addresses project to ODUG, which is rejected
- February 2014
  Department of Business, Innovation & Skills (BIS) publishes independent report by Hugh Neffendorf on feasibility of an open address gazetteer
- March 2014
  UK Government Public Administration Select Committee (PASC) publishes its report, Statistics and Open Data, which criticises the sale of PAF
- **May 2014**
  Open Data Institute submits amended proposal for funding to the Release of Data Fund, to develop Open Addresses
- **July 2014**
  Open Data Institute receives grant from Cabinet Office to support Discovery Phase of Open Addresses, and submits results from Discovery Phase
- **August 2014**
  Open Addresses Symposium held, bringing together open data community experts, addressing experts and organisations using or publishing open data
- **December 2014**
  Open Addresses Ltd receives grant from Cabinet Office to support Alpha and Beta phases of Open Addresses; Open Addresses Ltd comes out of dormancy
- **January 2015**
  Open Addresses launches photography competition, 'Picture an Address'
- **February 2015**
  Open Addresses attends and supports Open Data Camp 2015
- **March 2015**
  Release of Data Fund funding period ends; Ordnance Survey release new open datasets and lifts restrictions on use of UPRNs
- **April 2015**
  Open Addresses beta website launches; Open Addresses receives funding from BCS LISG Digital National Framework

What type of organisation should Open Addresses be?

One of the fundamental questions about the provision of an open address database is about the type of entity that should manage that publication. In this section we describe the Open Addresses organisation and the governance arrangements put in place around it. We then discuss more generally the kinds of attributes that entities need to maintain National Information Infrastructure, and assess how far different address suppliers – including Open Addresses – possess those attributes.

Open Addresses Ltd

After seeking legal advice, the Open Data Institute decided to establish a **wholly-owned subsidiary company** limited by shares (**Open Addresses Ltd**), to take forward the Open Addresses project.

Three factors influenced this decision:

1. The **risks** associated with Open Addresses, particularly the risk of litigation

2. The need to set up an entity **quickly, at a reasonable cost**

3. Uncertainty as to which short- and long-term **business models** would work for the project

It was crucial to ensure the ODI and its ongoing activities were protected from any liability for risks associated with the project. As a wholly-owned subsidiary company limited by shares, Open Addresses Ltd has a legal identity distinct from the ODI. This helps isolate the ODI from the risks associated with an Open Address Database, while affording the ODI a degree of control over how Open Addresses develops and the business models it adopts.

Alternative organisation types, such as [community interest companies](#) or charities, were also considered. However, they would have brought significant additional administrative burden and constrained the business models that could be explored. Converting a limited company into a community interest company or charity is simple, but the reverse is not.

The articles of association for Open Addresses Ltd similarly provide for maximum flexibility in the company's operation:

> *The objects of the Company are to research, develop, engage, facilitate, collaborate and commercialise in the field of open addresses.*

Alternative, more restricted articles of association were discussed by the Open Data Institute Board, but placing such limits were judged to be over-constrained and ultimately ineffective (as nothing would prevent further changes to the articles of association in the future).

To ensure independence of Open Addresses Ltd and to extend the understanding of the geospatial business environment, three unpaid non-executive directors were appointed following an open recruitment process:

- **Steven Feldman**, Geodata Consultant through KnowWhere Consulting
- **Andy Hird**, Managing Director of Aligned Assets Ltd
- **Mike Sanderson**, Director of Strategy at 1spatial

Jeni Tennison, Technical Director of the Open Data Institute, was appointed Executive Director. In addition, Heather Savory attended Board meetings as an Observer, in her capacity as the Chair of the Open Data User Group.

The ultimate **organisational structure** for Open Addresses is yet to be determined. Open Addresses Ltd could convert into a community interest company (CIC), if this is perceived as more trustworthy by potential customers and collaborators. But due to the administrative costs this would entail, it is only likely if Open Addresses proves sustainable over the medium-term.

The **public sector**, or an existing address supplier, could start to **provide open address data itself.** As an official source, and one likely to start with a database with a higher quality and quantity of addresses, this would be superior to Open Addresses. In this case such an organisation could adopt Open Addresses' intellectual assets, such as data and code, which are all published under an open licence. If it proved useful, other Open Addresses assets – such as the brand, domain and community – could also be transferred back to the public sector in support of its publication of open address data.

The evolution of a National Information Infrastructure

The Shakespeare Review of Public Sector Information (2013) highlighted the need for a framework to manage important data held by government. It recommended that, as part of a National Data Strategy, the UK Government define its National Core Reference Data.

> *We should define 'National Core Reference Data' as the most important data held by each government department and other publicly funded bodies; this should be identified by an external body; it should (a) identify and describe the key entities at the heart of a department's responsibilities and (b) form the foundation for a range of other datasets, both inside and outside government...*

The Cabinet Office has since led discussions of the **UK's National Information Infrastructure** (NII): the **UK Government's first attempt at a strategic plan for the governance, maintenance and accessibility of important data** held by government.

The Cabinet Office defines strategically important data as:

> *Data held by government which relates to the fundamental components of our society and is likely to have the broadest and most significant economic impact.*

The Open Data User Group (ODUG), who has been contributing to the NII discussions on behalf of the open data community, [views the NII](#) as comprising data that is open by default,

> *...maintained to agreed standards, which describes the fundamental components of our society and can be openly accessed, used and reused by anyone.*

Address data has always been considered a core element of the UK's NII. In its definition of National Core Reference Data, the Shakespeare Review [noted](#), '*we would expect to find the **connective tissue of place and location**, the administrative building blocks of registered legal entities, the details of land and property ownership.*'

But the Shakespeare Review, Cabinet Office and ODUG visions of a NII are explicitly concerned with data held by government. This approach doesn't capture how other important public infrastructures – beyond data – are owned and managed today.

If we think of a NII like other public infrastructure – roads, railways, energy supply, education – then different models of ownership exist. Today, some parts of the infrastructure are owned by government, some are owned by the private sector and operated under instructions from government, and some are operated entirely independently but still subject to regulation in various forms by government.

The privatisation of PAF demonstrates what is, in practice, a **mixed model for ownership of important public infrastructure**. The NII, like other infrastructure, will be managed by a range of organisation types. The government's framing of the NII – and the policies it creates to ensure NII data has broadest economic impact – should take into account the whole data ecosystem, not just those data assets it owns.

Similar considerations also apply at a local and a global level. At a local level, information infrastructure helps **make cities smarter** by informing citizens and decision makers. At a global level, it enables us to **monitor multinational organisations**, and agree on comparative statistics that help us **measure the progress of international policies**.

What is an information infrastructure?

Information infrastructure identifies data that underpins important services, products and research. It **helps society to function better** by making that data more accessible and better governed under a strategic framework.

Information Infrastructure is **built up of data from different groups –** this data can be closed, shared between specific organisations or openly licensed, and owned by governments, businesses and non-profits alike.

Information infrastructure helps us to recognise **data that has social, environmental and economic value,** and make that data as available and useable as possible to realise that value.

What's expected from an entity responsible for managing NII data?

What attributes should we expect in an entity who has responsibility for information that is fundamental to how society functions? There is no canonical list, but observations across a range of information holders – spanning government departments, non-profit organisations, private for-profit registries and charities – suggest some basic criteria that might be taken into account for the governance of any core data asset:

- **Long-term sustainability** – funding, governance support and purpose should be continuous. This doesn't necessarily mean the entity has to have existed for 50 years. Rather, that the way it's governed, its funding and its core purpose suggest it has stability and is able to manage the data in the long-term.
- **A perceived authority –** the entity needs to be considered a credible, authoritative source of the data it manages.
- **Transparency –** the entity needs to be transparent about where the data comes from and how it is processed. It should have mechanisms in place that enable stakeholders to ask questions about the data (ie for public sector organisations, via FOI laws).
- **Openness –** the entity should be open to feedback and updates to the data and responsive to requests from all users: both internal and external to the organisation, both direct and indirect customers, and for known and novel uses. It should be proactive in sharing data, code and processes with all stakeholders, and make the data as accessible as possible.
- **Commitment to the availability of data –** this is particularly important as the entity will usually hold a monopoly over that type of information, which confers the power to withhold access, distort competition, or neglect data quality. The entity should be incentivised to maintain and continue to provide access to high-quality data. This might be a financial incentive (revenue generating), a mandate to provide access (for example, via legislation), regulation, or a social purpose.
- **Agility –** in the data context, 'agility' has two meanings. The entity should be able to adapt its data management to a changing social and technological context, and meet evolving user needs. It should also be able to update the data quickly, including in response to external feedback.

Each of these attributes feed into an organisation that is trusted to maintain high quality data in the long-term, and supply it in ways needed by society. The attributes that are most desirable will vary based on the kind of data the entity manages, and how that data is used.

Information monopolies and maximising data's availability and use

The extent to how well different entities could maintain NII, based on their attributes at the time of writing this report, is summarised in the table below red indicates that the entity does not have the relevant attribute, green that it does, and orange that it partially satisfies that need.

All these assessments are mutable: it would be possible for Royal Mail to become more open, or for Open Addresses to achieve sustainability. The full details for these assessments are provided in Appendix 1.

How well could different entities maintain NII?

	Royal Mail	GeoPlace	Open Addresses
Sustainability	🟩	🟩	🟥
Authority	🟩	🟩	🟥
Transparency	🟧	🟩	🟩
Openness	🟧	🟧	🟩
Commitment to availability	🟧	🟧	🟧
Agility	🟥	🟧	🟧

It is natural for the maintainers of strategically important data to hold a monopoly position within the market – it is a side-effect of having an authoritative set of data and guaranteed longevity. The other attributes that we have highlighted act as a balance to that natural state, to increase trust in the entity by reusers of the data that they maintain.

Royal Mail is an established brand and longstanding provider of postal services for the UK. Similarly, while GeoPlace is a relatively new partnership, Ordnance Survey is the authority for high quality mapping data within the UK. Both are monopoly information providers, and so can define the market for their data and shape its use. They have a captive market, so there may be less incentive for them to adjust to changing user expectations, or respond quickly to feedback.

Royal Mail's litigious approach to the use of PAF by third parties has led to distrust within the data community regarding its mission and plans for the future availability of PAF. The costs it sets for use of PAF by individuals, government and companies cannot be seriously challenged, because there are no alternatives for users.

Similarly, the launch of GeoPlace has been met with criticism from the data community for allowing Ordnance Survey to expand its spatial address data monopoly. The data cooperation agreement between Ordnance Survey and local authorities behind GeoPlace compels them to provide the local authorities to give their address data to Ordnance Survey, who then sell it within the AddressBase product line. Even against a backdrop of Ordnance Survey's ongoing adoption of open data, there is little room for them to manoeuvre given that their data includes Royal Mail's and they are therefore limited by Royal Mail's terms.

The aim of an information infrastructure – defining a strategic approach, and identifying data that comprises it – is to make that data as widely available and useable as possible. As monopoly providers of UK address data in a weak regulatory atmosphere, GeoPlace and Royal Mail are able to prescribe what available and useable means.

Underlying the establishment of Open Addresses, and the support it has generated from the data community, is a concern that the existing providers of UK address data might not continue to improve the usability and availability of their data as user needs evolve, and societal expectations change.

Open Addresses attempts to challenge the monopoly these providers have over address data, via an open business model for address data. Its existence highlights the challenges of ensuring monopoly information providers make data that is part of an information infrastructure as available and useable as possible.

In many cases, entities responsible for strategically important data will be the monopoly providers of that information. The challenge is how to balance that position while ensuring that the data's accessibility and usability continues to evolve with user expectations.

Lessons learned and recommendations

We have to expect that our information infrastructure will be managed by different types of organisations, from private companies to the public sector. **But the government should play a role in ensuring that society can reliably access high quality data**, by:

1. **helping to sustain those managing the data**, either by supporting institutions that are already sustainable or by providing guaranteed, long-term support for new entities

2. **encouraging data managers to be transparent and open**, drawing on Freedom of Information laws or independent regulators

3. **mitigating the potential abuse of monopoly positions** by minimising the profit motive, for example by ensuring the entities are not for-profit companies or by limiting their ability to profit from the information assets they maintain

To be a convincing maintainer of UK addresses as NII, over the medium- to long-term, Open Addresses would need to:

1. **prove its sustainability**, either by coming under public ownership or demonstrating its ability to get sufficient revenue to cover its costs

2. **build an authoritative database of UK addresses**, reaching the same coverage as evidenced by Royal Mail and/or GeoPlace

3. **be subject to an institutional lock-in,** ensuring it provides free open address data, for example by becoming a community interest company with that express purpose

How can Open Addresses be sustainable?

For open data to be trustworthy, and therefore used, the organisations that publish it must have a sustainable business model that provides sufficient revenue to cover costs. These costs can include data maintenance, publication or provision through APIs, sales, accounting and legal expenses, and ongoing technical development.

One of the challenges for Open Addresses was finding a business model that could enable it to be sustaining in the long-term and (given that none of the Cabinet Office funding could be allocated after March 2015) survive in the medium-term. Over the course of the project, Open Addresses looked at the gaps in the current addressing market, which is described in Appendix 2, to identify a unique value proposition that could provide a sustainable revenue.

Open Addresses value proposition

For Open Addresses to have a sustainable business model, it had to identify a source of income. Given that the provision of open address data was at the heart of its proposition, the core approach adopted by Open Addresses was "**data for free, service for a fee**". Conceptually, this could take one of two forms:

1. **freemium access to data,** such that organisations who required a service level agreement that provided guaranteed levels of service would pay for those guarantees
2. **cross-subsidising the free provision of open data** by providing other address-related services that people would pay for

At this early stage in its evolution, Open Addresses cannot compete on quantity or quality with the address data provided by other suppliers. Address data from Royal Mail and Ordnance Survey has been built up through decades of investment from themselves, from government and from local authorities. And despite many open datasets containing address data, the legality of reusing those addresses was frequently called into question.

Open Addresses therefore needed to **identify a value proposition** that was sufficiently distinct to **compete in and grow the address data market**.

Keeping costs manageable

An obvious area on which Open Addresses could compete with other suppliers is on cost. Open Addresses has a **lower maintenance cost** because it has a collaborative maintenance model. It does not rely solely on postal workers or local authorities to supply address information, but accepts it from anyone who knows – be they companies, public authorities or individuals. This collaborative model is supported by the database being open: those who contribute addresses know that they will always be able to get hold of

(and freely reuse) the data that they have contributed – they are not simply contributing to the bottom line of another organisation but can themselves benefit from others contributions. The lower cost of maintaining the address database could be passed on to the consumers of address data.

Through conversations with existing users of address data, Open Addresses determined that customers for individual address lookup services, such as those used within websites, were generally happy with the cost of the services they were receiving. To bring these users into the collaborative network of contributors and users of Open Addresses data, Open Addresses would have to provide higher quality services.

This was not universally the case for the bulk users of address data. For example, Open Addresses talked to software suppliers who provide products that incorporate demographic and geospatial data. These suppliers explained that the majority of the cost of their product was due to the address data that they built in being expensive. If its cost were lower, such suppliers could be more profitable, either by having a greater margin on each sale or by expanding their customer base by offering a cheaper product.

In addition, while both Royal Mail and Ordnance Survey offer free access to address data for particular types of organisation, such as startups, some were put off by the costs that they anticipated incurring if they became more successful. This was particularly the case where the use of address data was not a core part of their business and therefore had less value to them.

Licensing openly for maximum use
There are two **advantages of open licensing** for reusers:

1. the licences themselves are **clear and easy to understand**, which makes them more accessible to those who do not have a legal background or access to lawyers
2. the terms of open licences provide a huge amount of **flexibility in what reusers can do with data**; with the CC-BY licence used by Open Addresses, so long as reusers provide attribution to the source of the data, they can do whatever they like with it

The complexity of the licensing arrangements for PAF was raised by Ofcom following its consultation in 2013. While Royal Mail licences have been subsequently simplified, the array of licensing options – with different permissions and different price points – are still hard for organisations to navigate, and the licences themselves run to pages of legal language. There are similar issues with Ordnance Survey licences. Even large companies are likely to find these off-putting.

Open Addresses cannot completely avoid legal complexity. Open Addresses needs to provide services that are sustainable, and therefore needs to enter into contractual agreements with its customers. However, **openly licensing data** drastically reduces the restrictions and uncertainty encountered by reusers as to whether they can use data to

do what they want to do. This can both **enable new applications, products and services** to be built on Open Addresses that would otherwise not be possible, and r**educe the time it takes for reusers to get started.**

Restrictions on what can be done with address data can hit even companies with plenty of financial reserves and legal expertise. Google, for example, does not use official address data (beyond postcode locations provided by CodePoint Open) within its maps of the UK, because of the licensing restrictions that prevent it from offering the services that benefit its users. As a result, Google has had to [construct its own address data](#) for the UK using the same labour-intensive methods that it employs in countries that lack official address data: methods that are completely **out of reach for other smaller organisations**.

Keeping content accurate and relevant

Address data within PAF focuses on locations to which mail can be delivered. The addresses within PAF are structured based on the routing of mail to those buildings: each address as a postal town to which mail is initially sent before being passed to postal workers to deliver on their rounds. This **postal town does not necessarily match the location of the address** or the affinity of the residents: homes in Wales may be assigned postal towns in England, if those towns are closer for the purpose of delivering mail.

AddressBase contains these same Royal Mail addresses, along with a location for each and a Unique Property Reference Number (UPRN) where possible. AddressBase Plus, on the other hand, also provides the addresses assigned by local authorities, which are sometimes different and more likely to reflect the physical location of the building, rather than its position within the mail delivery network. The address data maintained by local authorities also recognise more addresses than the delivery points listed in PAF, for example, where multiple properties share a single letter flap.

The **need for alternative addresses** arises in other situations too:

- **streets and localities in Wales may have both Welsh and English names** (eg *8 Hoel Croes y de* is the same as *8 Southcross Road*), so the preferred address will depend on the preferred language of whoever reads it
- **properties with names may also have a street number** (eg *Rose Cottage, Acacia Avenue* is the same as *8 Acacia Avenue*); the property owner may prefer the property to be addressed by name, while people looking for the property might find its number more useful
- **streets may have multiple names**; official street names listed by local authorities might not match the name-plates on the streets themselves, which are more useful for people looking to find them locally (and areused in community mapping projects such as OpenStreetMap)
- **some addresses have changed over time**, because their postcodes were recoded, for example, but older data may still list the older address

Open Addresses found that a single address (with one URL) should be presented in multiple formats, depending on how they are accessed and on user/resident language preferences.

Keeping data up-to-date and timely

The largest problem faced by address data users – both companies and end-users of address lookups that companies provide – is the lack of up-to-date address data provided by official sources. New versions of AddressBase are only provided [every six weeks](#). Changes to PAF can take [13 weeks](#), or even longer. These delays are exacerbated by organisations that host their own address lookup solutions but do not frequently update the address data that they use.

Delays in updates to address data impact new home owners and new or relocated businesses in particular. When businesses like [insurance companies rely on address data](#) to be able to issue cover, people can find it **hard to get the services that they need,** and those companies can likewise lose business. This also costs local authorities money when they have to intervene to ensure that new properties are habitable.

Open Addresses currently **generates a new bulk download every week** and incorporates new data into its API every day. The collaborative maintenance model means that people can add their new addresses directly into the database themselves – keeping administrative overheads very low. It mitigates against the problem of inaccurate addresses being added to the database by providing a **confidence score** against each address. It plans to provide three levels of downloads in the future: one that contains all addresses, one that contains only those in which it has a reasonable degree of confidence, and one in which it has high confidence. This gives address data users a choice based on how willing they are to accept the risk of inaccuracy.

Services to suit different datasets

For Open Addresses, new-build addresses are equivalent to addresses that it is not currently aware of, simply because it lacks data. Both types of addresses can be handled with a similar approach. Open Addresses can offer four types of services that do not require it to have a complete set of address data, and can be used to support new-build addresses:

1. **Parsing free text addresses into structured fields**; this can be informed by general gazetteers of streets and localities, which are available as open data, and does not require knowledge of every single address in the UK.
2. **Autocompleting addresses**, providing suggested street names and appropriate localities, can be done based on information from gazetteers supplemented by information about co-occurrence between streets, localities and postcodes within known address data.
3. **Providing confidence measures for addresses,** giving an indication of how likely it is that an address really exists, can be carried out based on similar building

blocks: gazetteers and statistical calculations based on co-incidence of address components.

4. **Matching incomplete or slightly different addresses** can be useful in cases where there are multiple sources of addresses, between an existing customer database and a newly completed form, for example; you can match two addresses without a full list of address data.

These services support **new kinds of interactions for websites** that require people to enter their addresses. Typically, sites offer separate fields for separate components of an address, allowing them to capture more structured data. Sometimes those separate fields are completed when a postcode is entered and address is selected from a drop-down list.

Based on user experience tests, Open Addresses found that free-form address entry – where people provide an address as they would on an envelope – are quicker and easier for people to complete than typical website forms. This is particularly true for less digitally-aware users, which is particularly important for government sites. To support this, addresses need to be intelligently parsed into their component parts. Confidence measures about the likelihood of an address existing are also helpful, so users can be prompted to correct addresses that they have accidentally misspelled, for example. Open Addresses services directly support better user experience when entering addresses on the web.

Importantly, each of these services also helps Open Addresses to **safely capture addresses without infringing third-party intellectual property rights**. They drive address submission into Open Addresses – boosting its coverage, increasing its quality and building up the value of the open address database – while providing a useful service to large and small businesses, charities and individuals.

Open Addresses business model

Open Addresses Ltd has not yet built a sustainable business.

The initial business model identified during the discovery phase mirrored those offered by existing suppliers and relied on freemium access to an open address database. While there is definitely a market for such a database, Open Addresses could not build one within the time available.

- When taking a risk-averse approach, as Open Addresses did, there is **very little available open address data** that can be used within an open address database.
- Tactics such as inference to increase coverage act as a multiplier on known data; a small initial set of addresses provides **limited scope for expansion** through inference.
- It costs time and money to build a community of contributors for such a project. **Legal safeguards** (such as the ability to take down addresses that infringed

third-party intellectual property rights) and **technologies** to facilitate contribution both take time to put in place.
- The unique value propositions of low-cost open data provided under an open licence were not valued highly enough to outweigh the **lack of quality and quantity** in the Open Addresses data.

However, having talked to potential customers, and acknowledged the lack of a large community of contributors (and the cost of developing one), Open Addresses has identified a **promising source of revenue** by providing address parsing, auto-completion, assessment and matching services using a freemium model, with:

- a **free tier** for light users of the services (eg < 50 requests / day)
- a **standard tier** for normal users (eg < 500 requests / day)
- an **unlimited tier** for enterprise users

Open Addresses also plans to offer a reduced price for users that agree that the addresses they submit are stored by Open Addresses. This reduced pricing, and the feel-good benefit of collaboratively creating a public good, should enable Open Addresses to **engage sufficient customers to fulfil its wider purpose of building a high-quality open address database**.

Open Addresses intends to initially offer its services for free, mainly so it can grow its customer base and see which services are most popular. To develop its services, Open Addresses is currently looking for and pursuing:

- **funded project and consultancy work** that can contribute to the long-term development of Open Addresses
- **additional funding** from donors and investors

The success of these investigations will determine whether Open Addresses Ltd succeeds in the long-term.

Open Addresses business model: lessons learned and recommendations

The current address data market is well-established, but there are several areas in which innovation could give address data users a better deal:

- Pricing models and licences for address data from existing suppliers are complex and geared towards pre-existing use models which can prevent innovative uses of address data. Free developer licenses do not sufficiently address the complexity of licensing for startups and small businesses.
- Address data needs to include more than house numbers, streets, localities and postcodes to be most useful; geolocations and addresses presented in different ways can be important.

- There are now a rich set of gazetteers available, which address-based services can be built on without necessarily accessing a full address database. These can be supplemented by statistical data from known addresses
- Creating a large enough customer base for a new collection of services to reach sustainable operation has not proved possible within the timeframe of the Open Addresses project.

What data can Open Addresses lawfully use?

Open Addresses needs to address **two areas of data-related law: intellectual property rights**, and **data protection and privacy issues**. This section explores the approaches taken by Open Addresses, and lessons learnt along the way.

Many of these issues are likely to arise in other initiatives, and can be applied and considered in broader open data projects.

Intellectual property rights

The goal of open data, and hence Open Addresses, is to enable as many people to reuse data as possible. Clear open licences assure reusers that they can use data for any purpose. But to license data, an organisation has to own it. Open Addresses has therefore worked hard to establish the ownership of address data. Appendix 3 describes the **rights situation around addresses** in detail, but the main conclusions are:

- There are **no intellectual property rights in a single address**, so individuals supplying single addresses do not need to explicitly reassign rights; however, a contributor agreement is necessary for significant contributions
- Address data that is validated or checked against existing address databases may be **contaminated with third-party intellectual property rights** and therefore cannot be reused by Open Addresses or similar efforts; this includes, with very few exceptions, addresses published within open data by the public sector under the Open Government Licence
- New addresses may be supplied to Open Addresses by local authorities, provided they are not accompanied by a geolocation derived from Ordnance Survey maps and do not contain information extracted from the Local Land and Property Gazetteers and supplied to GeoPlace

Data protection and privacy

The basic address data collected and collated by Open Addresses does not include any information about individuals. Open Addresses does not collect information about who lives at which address within its core data model.

However, the intention to record the sources of address data does lead to the possibility of provenance data revealing information about the people who live at an address. For example, if a utility supplier provides the addresses of their customers to Open Addresses, and the provenance record indicates that each relevant address was donated by that utility supplier, this reveals the choices that the residents have made about their

suppliers, as well as potentially commercially sensitive information about the supplier's customer base.

Open Addresses is therefore pursuing both a modified form of provenance information – in which the supplier of address data is indicated only through a non-disclosive category (such as "utility supplier") as well as a pseudonymous identifier whose mapping to true source is retained privately by Open Addresses.

If this proves to be an issue, Open Addresses also intends to set up an Ethics and Privacy Advisory Group to help understand and mitigate against any unintended impacts of its service making address data more widely available. We are not aware of any other address service provider that is working to mitigate the risks of holding information about the services used by those at a particular address.

Legal risks and mitigation

Operating in the addressing market comes with a high level of legal risk. The **monopoly providers of addressing data** are **protective of their market position**. This is evidenced by, for example, the [cease and desist letter](#) issued by Royal Mail in 2009 to Ernest Marples Postcodes Ltd, a UK postcode lookup service that provided web providers with an API to help people search for information specific to their area. It is also apparent in the ongoing [legal battle between Canada Post and Geocoder.ca](#) over a similar postcode lookup service based on crowd-sourced data.

The **challenges for small companies** operating in this area are threefold:

1. The **lawfulness of certain kinds of activities involving data is unclear** because it has not been tested in court.

2. Abiding by the law does not necessarily **prevent legal action being threatened or taken agains**t you by organisations with deep pockets.

3. The **cost of insurance against legal risk** in this area is out of reach for new businesses.

Open Addresses took a risk-averse stance. It invested in getting legal opinion about its actions. It put in place procedures and practices to ensure that it remained within the law.

But even with these mitigations in place, **Open Addresses was only able to find one insurer** who would provide it with cover for defence against Intellectual Property Infringement claims. Other insurers stated:

> "We believed that the Post Office's strategic interest in maintaining their exclusive ownership of the Postcode Address File would likely encourage them to test this uncertainty [in the use of the 'hosting' defence] in court."

The terms of the one insurer who could offer cover were unrealistic; they included an endorsement for Open Addresses to:

> "prior to hosting any data, ... undertake thorough checks of the source of data and has provisions in place to reject such data where it is derived from the Postcode Address File (PAF)"

It is generally recognised that content-hosting services, such as YouTube or Tumblr, cannot check every video or post that they host. The same is true for data; agreeing to this endorsement would make it impossible for Open Addresses to function as a platform for address data.

These costs are also currently out of reach for Open Addresses Ltd. The premium for this insurance was between £70k and £100k, to provide up to £5m cover with an excess of £35k plus 5-7.5% of the costs.

Unfortunately for Open Addresses, this situation creates FUD – [Fear, Uncertainty and Doubt](#) – on the part of prospective customers, partners and investors. If Open Addresses will need to stop publishing data, or even fold entirely, because it cannot stand up to legal action, stakeholders might always be uncertain about when that might occur, and whether to continue investing in it. The steps Open Addresses take to remain inside the law won't necessarily reduce this, as it stems from broader uncertainty around the law itself and the presumed litigious nature of Royal Mail.

Legal risks and mitigation: lessons learned and recommendations
Investigating third-party rights in data, particularly through FOI requests, is an involved, lengthy and costly process, both for potential reusers and for publishers responding to those requests. **Uncertainty over rights creates an environment where legal risk is high**, creating an atmosphere of fear, uncertainty and doubt that prevents innovative businesses from flourishing. We therefore recommend that:

- Efforts should be made to **clarify the legal situation regarding the infringement of database rights** when checking, validating or correcting data against another dataset. There is a lack of case law in the area which means it is unclear whether this is an infringing act.
- Royal Mail, Ordnance Survey and other organisations that encourage the use of their products for checking, validating and correcting data should **clarify rights over the results** of that process (eg Cleansed databases). Rights over the data should be transferred to the user in most cases.
- Customers of services that involve lookups, checking, validating and correcting data (including the UK public sector) should be aware of the **potential impact of introducing third-party rights into their data** – in particular, its impact on publishing open data for reuse.
- Data holders should, as part of their data asset management processes, **record the intellectual property rights associated** with the data they hold within their

data audit or information asset register. Within the public sector, this should be part of the responsibility of the Senior Information Risk Owner (SIRO).

- Publishers of open data should clarify who owns rights in the data, and highlight where third-party rights not covered by the licence exist in open data (and therefore are not licensed for reuse). It is also useful to explicitly state that no such third-party rights exist, where that is the case.

- Open data publishers should **publish versions of the open data** containing only intellectual property covered by the licence. It may be that a separate mark, or even a separate licence, would help to make this clearer for reusers.

- Where data is sourced from organisations, **provenance** for that data **may reveal information both about individuals and companies**; it must therefore be treated carefully.

What technical infrastructure does Open Addresses need?

Open Addresses has adopted several innovative approaches while working to supply open address data, provide services around that data, and facilitate collaborative maintenance of address data. All the source code developed for Open Addresses is available on GitHub, all licensed openly (most under the MIT licence), which means that it can be reused by anyone.

The components of the technical solution developed by Open Addresses is shown in the diagram below:
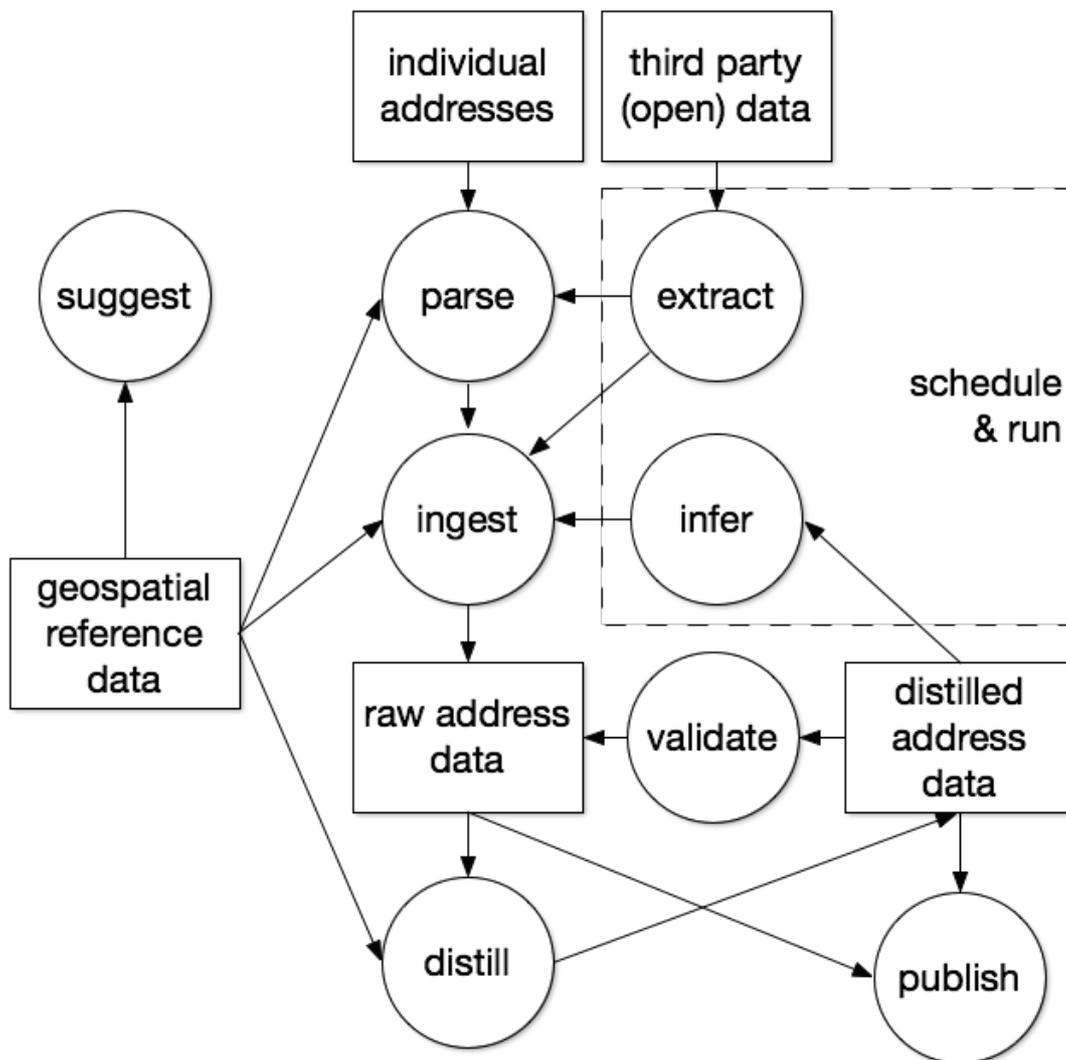
Each of these components is described within this section.

Parse

The simplest entry point to the Open Addresses system is the address parser, sorting office, which runs as a web service or API. The parser takes addresses provided simply as a string and uses data about postcodes, localities, streets and so on to parse the address into its components. The structure of the address generated through parsing, and used throughout the system, currently mirrors the BS7666 standard, with components for:

- saon - used for units or flats
- paon - used for building names or numbers
- street
- town
- postcode

Third parties can embed calls to the parser into their websites or use it as a service within scripts. There is an optional switch to determine whether or not addresses that are run through the parser are also submitted to Open Addresses. It is possible to use the service to parse addresses without those addresses then making it into the data stored by Open Addresses.

Ingest

The ingester, Ernest, takes parsed address data and the provenance for that address data, assigns a confidence to that address, and records it as raw address data.

Recording provenance

It is important for Open Addresses to **record the provenance of each address** that it holds, for two main reasons:

1. it helps to **promote trust** in the data by its users, because they can see where the data came from; in particular it highlights that the sources of the data are not tainted with third-party intellectual property rights
2. it helps Open Addresses to **trace and remove data** from contaminated sources, ensuring that they are able to respond to take-down notices quickly, or simply respond to trace bugs in extraction scripts

While these reasons are both particularly salient for Open Addresses – because of the potentially hostile legal atmosphere in which it exists – this also applies in other situations where data is sourced from several places: OpenCorporates, for example, records and makes available provenance for each statement that it holds.

Ingestion is always invoked through a script of some kind. The **provenance data model** used by during ingestion by Open Addresses includes:

- a **URL** for the source of the data
- the **string** that was the original user input for the address
- a **timestamp** for when the script processing that data was run
- a **reference** to the version of script that generated the address data
- a **free-form attribution string** which may be used to hold the name of the source

Calculating confidence

When data appears within a database, it is usually assumed to be 100% correct, even if that assumption is unfounded.

The "ground truth" of addresses changes all the time. New buildings are built; old buildings are destroyed, divided and combined. Streets are constructed, renamed and diverted. Postcode assignments are changed to suit the logistics of mail delivery. Equally, addresses may be entered into systems with typos, or address parsing or inference may go wrong, generating completely spurious address data.

Open Addresses allocates a confidence score to each address, so address data users can understand the risk of a given address being wrong.

Confidence is calculated during the ingestion and distillation stages. During ingestion, the ingestor calculates a confidence score for that address. The confidence score is a measure of how likely it is that the address provided actually exists in the real world. To work this out, Open Addresses combines:

1. the likelihood that an address in the specified **postcode** sector is also in the specified **town**
2. the likelihood that an address in the specified **postcode** is in the specified **street**

This calculated confidence is then adjusted based on the date at which the address was entered in the original data – the date that a company address was last updated, for example. This gives a final confidence score for each submitted address, which can range from 0 to 1000.

Distill

Raw address data contains a single entry for each address ingested into Open Addresses, along with the provenance of that address. A single real-world address may appear many times within the raw address data, if it has been encountered many times within the original data.

The distillation process matches up multiple occurrences of the same address to create a single entry in the distilled address data for each. This includes combining provenance and confidence information from the addresses.

Calculating confidence

During distillation, the confidence level for an address is adjusted to incorporate a judgement about the reliability of the sources of information about that address. For example, an address supplied by a local authority will have a higher confidence level than one entered submitted anonymously on the Open Addresses website. Confidence will be even higher if the same address was submitted through multiple sources.

Publish

Both raw and distilled address data is published on the Open Addresses website. Bulk downloads containing all address data are currently generated on a weekly basis:

- in CSV and JSON format
- as large single files or split into separate files by postcode sector
- through a normal HTTP download or through BitTorrent

The use of BitTorrent to enable downloads reduces transmission costs for Open Addresses for what should grow into a large file. Unfortunately, torrent clients on some operating systems are hard to use and may be blocked, so standard HTTP downloads are also offered.

Each address in the distilled data is also given a URL, so that there is a web page for each address, such as the one shown here:

The URLs for addresses provide an opportunity to link other data about an address together, using a single identifier. While the Unique Property Reference Numbers (UPRNs) provided by GeoPlace could be good identifiers for addresses, these are not available in bulk and therefore Open Addresses cannot rely on them for address identity.

The Open Addresses website allows for address search based on street, town and postcode.

Extract

One way to bring addresses into Open Addresses is by extracting them from other datasets. Open Addresses uses Turbot, a sandboxed environment (originally developed by OpenCorporates) to enable third parties to create extraction scripts and run them periodically.

Third parties can write scripts in Ruby, create a manifest file that describes the script, and register them with Turbot. Registered scripts must be available under an open licence and visible for review by the Open Addresses team. If they are approved by the team, the scripts are scheduled and run on a routine basis – for example, the Companies House extraction script runs monthly as the data is published monthly.

While this environment is intended to enable collaboration on extracting address data, so far there have not been any third party script registrations. This might be due to target datasets being hard to identify, third parties lacking motivation, or barriers in the process itself.
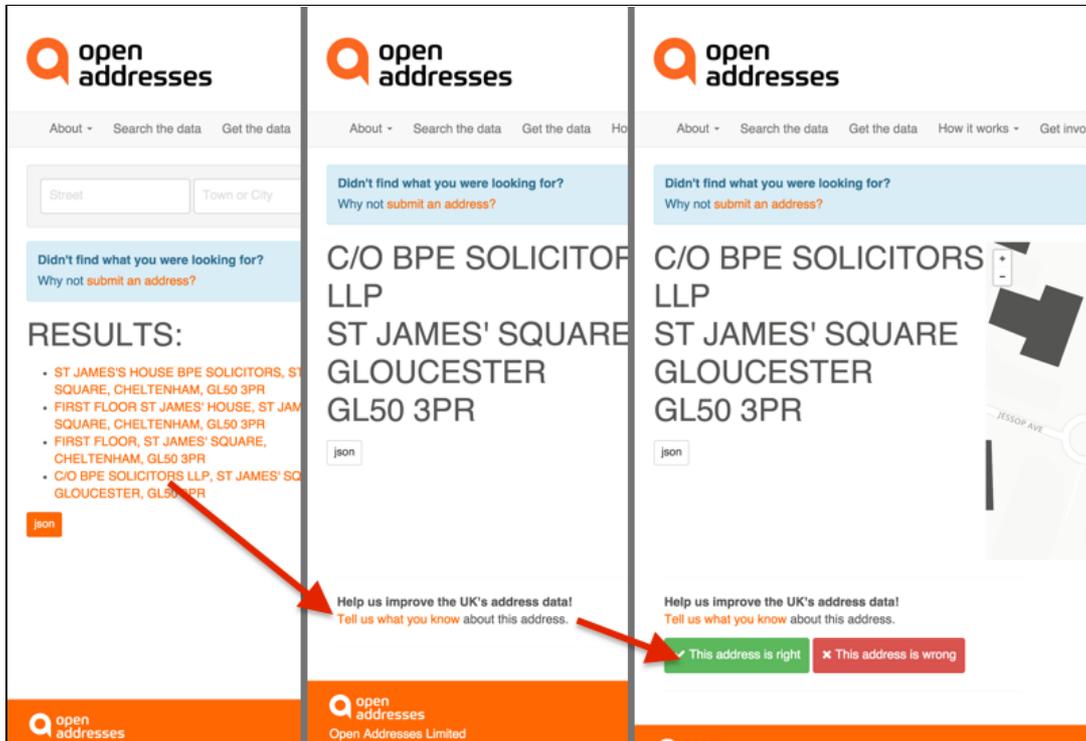
Infer

One way of boosting the limited amount of address data that is available for reuse is to infer addresses based on known addresses. For example, if there is data about 5 High Street and 30 High Street, Open Addresses will infer – with low confidence and clearly labelled as inferred addresses – that there may be addresses between those two numbers as well.

Open Addresses has only implemented this basic form of inference. More sophisticated inference could use information from maps about the number and locations of buildings in a street to supplement inference algorithms.

Validate

Because addresses within Open Addresses may come from sources that are out of date untrustworthy, or be incorrectly inferred, Open Addresses provides an additional API that enables people to indicate whether an address does exist or not. This API can be accessed through the website – as shown below – or used by other products.

Suggest

The fundamental geographic building blocks that are used when parsing, inferring and distilling addresses can also be used to suggest address elements that contain particular strings. This can be used as the basis for auto-completion services that help users to enter addresses in websites.

> **Building infrastructure: lessons learned and recommendations**
> Open Addresses has experimented with several novel approaches to help it maintain and publish open data. They have found:
>
> - Tracking and reporting the provenance of data takes **time and technical expertise**. Projects that do this must understand and justify the investment. For Open Addresses this was legal protection but in other circumstances it might be increased trust or the provision of attribution to contributors.
> - **Calculating and providing the confidence for data** is a novel way of enabling anyone to contribute to the dataset without having a naive assumption that all contributions will be correct. This approach should be explored further in other **environments that require collaborative data maintenance**, along with fast, non-memory-intensive calculation techniques.
> - **Sharing data over BitTorrent promises lower costs** for data holders but has the disadvantage of **raising access barriers for potential reusers**. It also

makes it hard to quickly retract information, if that data has been distributed across other computers for download.

- Bulk ingestion of third party data has promise for other collaborative data projects, but only if developers are motivated to contribute by **identifying suitable sources and writing the code required for ingestion**. For Open Addresses, other contribution methods show more promise.

Conclusions

The Open Addresses project aimed to investigate whether it was viable to **build and maintain a sustainable open address database using collaboration, cross-subsidy through services and volunteer effort**, or whether the only route to open address data within the UK was continued lobbying to open existing sources of address data owned by Royal Mail and GeoPlace.

The conclusion at this stage is mixed. It is clear that Open Addresses has not managed, within its 5 months of operation, to create an authoritative database of UK addresses. Nor has Open Addresses Ltd built, within that time, a sufficient collaborator or customer base to enable it to sustain itself.

On the other hand, the project has **identified a promising set of address-based services** that could lead to both sustainability and the creation of a high-quality authoritative set of address data in the long term. It has constructed and contributed to a **useful set of open source tools**. It has exposed a number of **wider lessons that could be picked up by existing address data maintainers**, the developers of future open data projects, and by policy officials framing open data strategy for the UK.

Open Addresses has found a real **hunger for better access to address data** within the UK. This hunger arises from:

- a wish for **simpler and cheaper pricing** structures
- a desire for simpler and **more flexible licensing** arrangements
- a preference for support for **multiple address forms**
- a need for a much **shorter update cycle**

It would be good to see more address suppliers move in these directions, to satisfy the needs of data users.

Open Addresses has also tested the approach of **charging for services rather than data**. This is a well-established pattern in software development, where organisations commonly charge for services built on open source code. It **engenders trust, promotes collaboration, and ensures a level playing field** rather than a false competition between a monopoly supplier and service providers, who are both competitors and customers.

Open Addresses identified **four services** for which there is a gap in the market, and where submitted addresses could help maintain an address database:

- address parsing
- auto-completion
- confidence scoring
- address matching

The Open Addresses project raises important questions about the governance of National Information Infrastructure assets. It highlights characteristics that entities managing this

type of information need to have in order to invoke trust in the ongoing availability of high-quality data, namely:

- longevity
- authority
- transparency
- openness
- commitment to the availability of data
- agility

It is possible for these attributes to be found within commercial organisations but because entities that manage information infrastructure are natural monopolies, checks and balances are necessary to prevent distortion of the market. Open Addresses does not yet meet all of these criteria: it needs to be sustainable, grow its database, and be constrained in how it can change in future.

Finally, Open Addresses has highlighted **the impact of legal uncertainties** that surround:

- the reuse of data published as open data (but that contain third-party intellectual property)
- the rights embedded in data that is validated against a third party's database, where those rights are not explicit conferred
- the applicability of a 'hosting' defence to crowd-sourced databases

Without greater legal certainty in these areas, innovation in the provision of data-based services and products will be limited to large companies that can afford the potential consequences.

At this stage, **the future for open addressing in the UK is unclear**. Open Addresses requires more investment and more time to reach sustainability. Meanwhile, the government's own needs for services around address data are becoming more salient. The political landscape may be about to change following the general election, and the impact of Ordnance Survey's transition to a government-owned company may start to be felt.

One thing is certain: the need for open address data as a fundamental building block in the UK's National Information Infrastructure will only grow over the coming years.

Appendix 1: The governance of UK addresses

Each of the existing **major address data providers in the UK** – Royal Mail and GeoPlace – can be evaluated against desirable attributes for an organisation **maintaining National Information Infrastructure**. As can Open Addresses. It is worth considering Royal Mail and GeoPlace, as they are both authoritative suppliers of address data in the UK.

Long-term sustainability

Royal Mail was established in 1516, and is the main longstanding supplier of postal services for Great Britain. Now privately owned, it is governed by the *Postal Services Act 2011*. Royal Mail's sustainability is provided through its main business line: postal delivery.

GeoPlace, a public sector limited liability partnership between Local Government Association and **Ordnance Survey,** was only established in 2010. GeoPlace is a data cooperation arrangement that requires local authorities in England & Wales to supply Ordnance Survey with address data, for which Ordnance Survey then sells licenses. While GeoPlace is young, Ordnance Survey was established in 1791, and is the national mapping agency for the UK as well as the world's largest producer of maps. Ordnance Survey's sustainability is provided through the commercial and public sector revenue it receives for the range of services that it offers, which includes the sale of address data.

In contrast, **Open Addresses** was only established in 2014. The development of Open Addresses has been funded by the UK Cabinet Office, with this funding ending in May 2015. It does not have the longevity of Royal Mail or Ordnance Survey, both in terms of its years in operation and guarantees of sustained funding.

A perceived authority

Royal Mail is the authoritative UK source of data about delivery points or postal addresses, with the Postcode Address File (PAF) containing over 1.8 million postcodes and 28 million addresses. It also assigns postcodes, and therefore is the authoritative source for **which postcode applies to which address**. PAF is oriented towards the delivery of mail. It doesn't contain the addresses for all properties (for example, a block of flats might have only one letter box, but 12 flats).

GeoPlace, the national gazetteer for spatial address and street information management, supplements this data, bringing together local authorities data on addresses used, for example, for council tax billing. It manages over 40 million addresses, and is considered authoritative for **many public sector purposes**, such as the census or emergency response.

While **Open Addresses** aspires to become an **authoritative source for UK address data**, it is still in beta phase and contains only just over 1 million addresses. Its address database could not yet be considered authoritative.

Transparency

As a private company, **Royal Mail** is no longer subject to FOI laws. It does report to the PAF Advisory Board on its management of PAF. The minutes from those meetings are available online. There is **little transparency about the way in which PAF is maintained**.

GeoPlace has responded to FOI requests in the past but it's unclear whether it is obliged to do so under FOI laws. While **Ordnance Survey** transitioned to a government-owned company in April 2015, it indicates it will still consider itself bound by FOI laws. GeoPlace is also **transparent about the processes, formats and agreements** that are involved in maintaining data.

Open Addresses is not subject to FOI laws. However, it has instigated a number of transparency mechanisms. Its use of a Github repository for management of the code behind the address database enables **transparent tracking of updates to the code, issues that have been raised and how these have been addressed**. The data itself also contains provenance information that provides **transparency about its source**.

Openness

Companies and individuals can make requests for changes to PAF via **Royal Mail** and via local authorities. The PAF Advisory Board provides an official route for requesting changes and making submissions for stakeholders to influence the terms under which PAF is made available. However, [Royal Mail has recognised](#) the **limits to its responsiveness to external feedback**, with users encountering delays between changes being requested and implemented.

Updates to the data held by **GeoPlace** are generally made via requests to local authorities. Responsiveness to those requests **depends largely on the local authority**.

Openness is core to **Open Addresses**' mission. Its address database can be updated by anyone. Feedback and issues are managed via its Github repository, user forums and social media presence. Its [mission statement](#) includes a **commitment to open by default, including open source code and roadmap**.

Commitment to the availability of data

Royal Mail's roles and therefore motivations are mixed. It is a creator and maintainer of address data and provides address data to service aggregators and providers. It is also a competitor to those same service providers by the direct sale of both address data and services. It also consumes address data through its postal delivery services. Royal Mail is in the unique position of having **early access to address data** and being able to ensure that postcodes are allocated in a way that suits its needs.

However, Royal Mail is regulated: it is required to maintain PAF and make it available on 'reasonable terms' under the [Postal Services Act (2000) s. 116](). It is subject to regulation by OFCOM. Royal Mail also generates revenue from PAF, which acts as an incentive to continue to maintain and make the data available. What constitutes 'reasonable terms' for PAF has been debated, and the limitations of the voluntary profit cap were recognised in the 2013 [OFCOM consultation](). **Current pricing mechanisms have been criticised as complex and administratively expensive**.

GeoPlace makes its data available through **Ordnance Survey**. For Ordnance Survey, address data is a source of revenue, which provides a strong incentive to continue to ensure that it is updated and to provide access to it. How Ordnance Survey provides access to address data is to an extent prescribed by the Re-use of Public Sector Information Regulations. Ordnance Survey became a government-owned company in April 2015 but [remains bound by those regulations]() and the Information Fair Trader Scheme (IFTS). However, the efficacy of the current regulatory framework [has been criticised]() and it is not yet clear whether it will improve in the future. Changing business models may impact on how Ordnance Survey collects, manages and provides access to address data in future. Finally, because GeoPlace's address data includes Royal Mail data, there are **limits to the flexibility that Ordnance Survey has to make the data available** under different terms.

Open Addresses is mission-driven. At its core is a commitment to provide free access to an open UK national address database. Open Addresses is a Limited Company governed by shares, but its sole shareholder is the Open Data Institute, which is itself a **mission-driven non-profit committed to growing open data culture**. On the other hand, Open Addresses is not subject to any regulatory oversight nor bound to its purpose by articles of association or similar mechanisms.

Agility

Royal Mail's ability to adapt quickly to new technologies and evolving user needs may be impeded by legacy systems and means of collecting address data, as well its size. Their position as a monopoly information provider may disincentivise agility in response to user feedback or new technologies. As the source of postcodes and the data provider most

others rely on (including GeoPlace), they have a **captive market**, and licensees of PAF are obliged to accept PAF standards.

The monopoly position of **GeoPlace** and **Ordnance Survey** may also disincentivise agility in response to user feedback or new technologies. However, the conversion of Ordnance Survey into a government-owned company was in part intended to promote a more agile and commercial stance.

As a young, small company **Open Addresses** is likely to be able to exercise greater agility in response to technological changes. It has established mechanisms for rapid responses to user feedback, via platforms like Github. Its desire to establish itself as a viable competitor incentivises agility. However, its ability to respond to evolving user needs and new technologies may be curtailed by limited funds and resources as it grows its business model.

Appendix 2: The current addressing market

The current addressing market in the UK is dominated by **two main sources of address data**: **Royal Mail** and **Ordnance Survey** as the delivery vehicle for **GeoPlace**'s data. Service aggregators and resellers provide access to these underlying datasets under a variety of terms and pricing models.

There are also alternative approaches to providing address information within the UK, such as identifying locations through their geolocation rather than their street address.

Royal Mail

Royal Mail offer a range of addressing products under the [Powered by PAF](#) brand name. This product set contains approximately 29 million addressable locations and has approximately 37,000 direct and indirect customers. Indirect customers are served by a network of resellers and service aggregators. The Powered by PAF site allows organisations who wish to use addresses to [search](#) for both direct and indirectly supplied products according to the use to which they wish to put the data.

The PAF product range does not contain spatial coordinates and no fee is payable to Ordnance Survey. Some resellers aggregate Royal Mail PAF data with OS and other sources.

Direct PAF pricing varies by the use to which the data is put. Royal Mail provides free access to small independent charitable organisations and for one year to independent micro-businesses who are developing PAF-based products.

The basic Address Finder and Manager products supply a simple bulk list on CD at a cost of £235+VAT but if the user wishes to have a set of utilities to help them use the data then an extra £750+VAT is charged. If an organisation wishes to resell the data then a charge is made for offering a bureau service, whilst if an organisation wants to have multiple users, or "seats", simultaneously using the data then again extra charges apply. The basic Address Finder product for 101 users will cost £16,985+VAT per annum.

During our discussions with potential users of Open Addresses data it was suggested that a range of new products might be offered by Royal Mail, for example, APIs to support predictive text entry capabilities for online services. However, no details for these services appear to be publicly available so the information cannot be fully verified.

Ordnance Survey

Ordnance Survey's addressing products are supplied under the AddressBase brand name. There are three products:

- **AddressBase** provides the same addresses as in PAF with spatial coordinates and a UPRN where applicable
- **AddressBase Plus** in addition includes addresses maintained by local authorities and multiple-occupancy addresses
- **AddressBase Premium** in addition includes historic addresses and alternative addresses

These products are supplied on CD, DVD or via FTP (for datasets less than 2Gb) and are refreshed on a 6-weekly period.

Customers must pay both a fee to Ordnance Survey and a fee to Royal Mail for using AddressBase products. A basic version of AddressBase for the whole UK for use by 101 users costs £129,950+VAT in OS fees and £4,200+VAT in Royal Mail fees (see the [full price list](#)).

A large number of public sector organisations have access to Royal Mail and Ordnance Survey addressing products under the Public Sector Mapping Agreement (PSMA). The pricing and terms for the PSMA and for the public sector PAF licence is negotiated by the department for Business Industry & Skills (BIS).

Service aggregators and resellers

Several organisations resell Ordnance Survey and/or Royal Mail data; some aggregate this data with other datasets for additional value. Examples of these organisations are:

- [PostcodeAnywhere](#) - which offers a range of API services such as address entry, data cleanse or geocoding priced by volume or lookup.
- [GB Group](#) - which offers both standalone addressing services (Matchcode Names) and aggregated services such as identity verification.
- [Experian QAS](#) - which aggregates address data with demographic data and personal data for applications such as targeted marketing.

The relationship between third-party service providers and the main sources of data (Royal Mail and Ordnance Survey) is sometimes problematic. Royal Mail and Ordnance Survey can be seen as competitors, particularly as they try to expand their service offerings into markets that service providers currently occupy. But they are also suppliers to those service providers, exacting a charge for the use of the data that they maintain.

Emerging location providers

Addresses (house numbers, road names and postcodes) developed as a way to determine a location based on the physical infrastructure of streets and buildings.

Locations can also be identified through latitude and longitude, but these numeric location identifiers are difficult for humans to remember accurately. In recent years, memorable encodings of geolocations have been created. These have the benefit of providing locations for things that have no visible physical infrastructure (such as a particular point in a field) and providing a global solution, of particular benefit in places where addressing infrastructure such as street names is partial or non-existent.

Whilst neither of the providers listed here have significant market share (in fact one was only launched in April 2015), long-term global approaches to providing locations may not involve traditional addresses at all.

- What3Words divides the world into a set of 3m by 3m squares each of which is allocated a unique combination of three words. What3Words is currently free for non-commercial use. Its licensing terms do not allow users to "*pre-fetch, cache, index, copy, re-utilise, extract or store any what3words Data*".
- Mapscode is run as a not-for-profit organisation, originally set up by TomTom, which defines a method of generating a short alpha-numeric string for each location.
- Open Location Code was announced by Google in April 2015. The codes are free to get and free to decode with open source libraries available on GitHub.

Appendix 3: Intellectual property rights in address data

There are two types of intellectual property rights that need to be considered in the UK: copyright and database rights. Copyright applies to works (some examples being photographs, musical compositions, or poems), whereas database rights apply to structured data.

Copyright in address data

Copyright applies to works that are created through original intellectual effort. In addressing, there are three areas where copyright could be pertinent:

- the name of a street or locality
- a postcode
- the selection or arrangement of database contents (eg of PAF or the NAG)

Street and locality names are typically generated either by property developers or by a local authority. Postcodes, which are an encoding of routing information for mail delivered by Royal Mail, are assigned to addresses by Royal Mail (though frequently postcodes for new properties are suggested by property developers or the local authority).

Legal advice received by ODI during the discovery phase for Open Addresses indicated that there is a low risk of copyright infringement in the use of postcodes or street and locality names within address data. In particular, it seems unlikely that the creators of street or locality names would pursue people who use those names for copyright infringement, and unlikely that there is copyright in the creation of a postcode.

Open Addresses did not replicate the selection or arrangement of existing address databases, so the issue of copyright in them is irrelevant for Open Addresses' purposes.

Database rights in address databases

Database rights apply to a collection of structured data that the creator has put substantial effort into obtaining, verifying or presenting.

There are many organisations that maintain address databases for the UK and therefore have database rights over those databases. Royal Mail is obliged by law to maintain the Postcode Address File (PAF) and make it available for reuse. GeoPlace maintains the National Address Gazetteer (NAG) which Ordnance Survey provides through the AddressBase suite of products. Local authorities maintain their own address data for their area. Google maintains address data for Google Maps which it makes available through the Google Maps API. Other companies have typically either built up address databases from their customers or supplemented old versions of PAF or AddressBase.

The activities that infringe database rights are defined in [regulation 16 of The Copyright and Rights in Databases Regulations 1997](#):

Acts infringing database right

16.—(1) Subject to the provisions of this Part, a person infringes database right in a database if, without the consent of the owner of the right, he extracts or re-utilises all or a substantial part of the contents of the database.

(2) For the purposes of this Part, the repeated and systematic extraction or re-utilisation of insubstantial parts of the contents of a database may amount to the extraction or re-utilisation of a substantial part of those contents.

Copying all or a substantial proportion of an existing address database directly, without permission, would constitute an infringement of database rights. Similarly, creating an address database by repeatedly querying [Royal Mail's address finder](#) or the [Google Geocoding API](#) would infringe any database rights held by Royal Mail or Google (as well as violating the terms and conditions for those services).

For open address data, there are three particular circumstances of interest, reflecting the different ways in which addresses can be supplied to Open Addresses.

Accepting single addresses

A single address, such as an individual might enter through a form, does not count as a database and therefore does not attract database rights and does not require intellectual effort to create and therefore does not contain copyright. This contrasts with submissions to Wikipedia, where much of the textual content submitted by contributors is created through intellectual effort and therefore contains copyright. Therefore unlike Wikipedia, Open Addresses does not require people who supply individual addresses to agree to a Contributor Licence Agreement.

Accepting addresses from local authorities

Local authorities maintain Local Land and Property Gazetteers (LLPG) that incorporate data from GeoPlace and Royal Mail, and from which data is supplied to GeoPlace and Royal Mail. They also maintain data about new addresses within local systems while in negotiation with property developers, prior to it being entered into their LLPG.

The rights over the data supplied to GeoPlace by local authorities (known as Authority Updates) and to local authorities by GeoPlace (known as Supplied Data). Participating local authorities have signed the [Data Co-operation Agreement](#) which includes the following clauses:

9   **Intellectual property rights**

9.1   As between the parties to this Agreement, the Authority owns and shall continue to own all IPR in the Authority Owned Data. Without prejudice to clause 3.1, this Agreement does not transfer any of these rights to GeoPlace.

9.2   Subject to clause 9.1, GeoPlace (or, where applicable, GeoPlace's suppliers) owns and shall continue to own all IPR in the Supplied Data and the Authority Updates, and the Authority's use or possession of any Supplied Data or Authority Updates does not give the Authority any ownership of or any interest in any of the Supplied Data or Authority Updates. In particular, the Authority acknowledges that Ordnance Survey has expressly reserved and retained all IPR in the OS Data and in any copies or adaptations or derived products made by GeoPlace or LGIH or the Authority (including in the GeoPlace Databases).

9.3   Where the Authority suspects any infringement or any other breach by a third party of any Intellectual Property Rights in any Supplied Data or Authority Owned Data used in the Authority Updates, the Authority shall notify GeoPlace and give GeoPlace and/or its third party licensors (including Ordnance Survey, Royal Mail Group plc and LGIH) all reasonably required assistance in pursuing any infringement.

This makes clear that rights in data obtained from GeoPlace are not transferred to local authorities, but that rights in data sent to GeoPlace are transferred to GeoPlace.

However, this rights assignment only applies to Authority Updates, not to the data the local authority maintains itself. Thus **local authorities can publish data that they have gathered independently (not from GeoPlace) as open data**, or supply it to Open Addresses directly. While they cannot publish data obtained from GeoPlace (such as data within their LLPG), [changes introduced in March 2015 to Ordnance Survey's approach to derived data](#) means that it may be possible for local authorities to also publish the locations of new addresses derived from OS maps.

Extracting addresses from OpenStreetMap

[OpenStreetMap](#) contains approximately 800,000 UK addresses gathered by the OpenStreetMap community through a combination of bulk data and street-level crowdsourcing.

The addresses are available as open data under the [Open Data Commons Open Database Licence (ODbL)](#), which is a share-alike licence that requires reusers of the data to release derived data under the same licence. This licence is incompatible with the CC-BY-4.0 (attribution only) licence chosen by Open Addresses UK, so it is not possible to import these addresses into the Open Addresses platform.

Contributors wishing to contribute into both OpenStreetMap and Open Addresses can do so, as long as none of the data that is submitted to Open Addresses (such as street names or localities) originally came from OpenStreetMap.

The OpenStreetMap community can import and republish the Open Addresses data as long as attribution is retained.

Extracting addresses from public sector open data

Many datasets published by the public sector as open data contain addresses. Large examples include Companies House data, the National Register of Social Housing (NROSH), Land Registry data, and data from the VOA. These datasets are all published or otherwise made available under the [UK Open Government Licence](.).

However, in common with other open licences, the Open Government Licence "does not cover: ... third party rights the Information Provider is not authorised to license". The impact of this exception is one of the main lessons learned from Open Addresses. It is not common practice for publishers of open data to explicitly state whether there are unlicensed third-party rights in the data and without these explicit statements, it can be hard to tell.

To use addresses from open datasets, Open Addresses needed to check whether that data included data with third-party database rights; these could arise:

1. if the data was directly copied from an existing address database

2. if an existing list of addresses (obtained through another route) was corrected or validated based on an existing address database

The public sector is licensed to cleanse its data (the second of these) using PAF under the terms of the [PAF Public Sector Licence](.) which is part of the Public Sector Mapping Agreement. This licence states:

> **2. *Database cleansing***
>
>   You may only provide access to Cleansed databases to third parties where:
>
>   (a)   such supply is not related to a service comprising the Database Cleansing of a third party's database and the supply of the resulting Cleansed third party's database back to them
>
>   (b)   if such databases are Substantially All Databases:
>
>   (i)   such databases are not represented or held out as a master, original or comprehensive address database or other similar description
>
>   (ii)   the access is provided in the course of your Core Business and is not carried on as a business in its own right, and

This makes it clear that organisations can provide access to cleansed data (so they are not violating the terms of the licence by publishing open data containing cleansed addresses). However, this does not provide such organisations with rights to the cleansed data.

To avoid the risk of infringing Royal Mail's database rights by taking addresses from public sector open data, Open Addresses therefore contacted each of the publishers of open data containing addresses, through a public Freedom of Information request, to try to ascertain whether the data contained third-party rights. The responses from the four large datasets for which FOIs were originally issued were:

- **Companies House** do not validate or correct addresses, meaning that addresses in their data could be reused
- **Land Registry** use AddressBase extensively, meaning that addresses in their data could not be reused
- **NROSH** was created by aggregating data from a number of different sources, some of which may have been validated or corrected (ie it is not clear what rights reside in the data)
- **VOA** could not respond within the time assigned for response to FOI requests (ie they also did not have the information to hand)

Erring on the side of caution, Open Addresses used only the Companies House data in the alpha and beta version of the site, giving 1.7m addresses. We note however that the introduction of new systems within Companies House may mean that this data also becomes unusable for addresses shortly. Open Addresses maintain a [Trello board](#) that lists other potential datasets and facilitates collaboration over investigating whether they can be used or not.

The following table shows the potential number of addresses that could have been used from key other sources, if a less risk averse strategy had been adopted. Note that none of these sources cover Scotland or Northern Ireland and they are likely to overlap. Being able to use all sources would have provided over 80% of addresses in England and Wales.

Source	Number of addresses
Valuation Office Agency	27 million
Edited Electoral Roll	13 million
Land Registry Price Paid Data	11 million
NROSH	4.5 million

Receiving addresses from third parties

It is impractical for a platform like Open Addresses that supports collaborative maintenance of data to check the intellectual property rights within data that is supplied by third parties. Open Addresses is similar to YouTube or Wikipedia in this regard: it cannot police what people provide, but it can ensure that it discourages rights infringement and enables rights holders to remove infringing material.

The submission guidelines for Open Addresses discourage the provision of data owned by third parties (for example through a script):

> Please be sure not to violate others' intellectual property or privacy rights with your submission. If you think your Intellectual Property rights are being infringed then please report the infringement.

> By submitting an address you are saying that Open Addresses Limited can re-use the address and publish it under an open data licence. You are saying that the publication of the address by Open Addresses Limited or third parties does not and will not infringe any of your legal rights or those of any third party.

The Trello board of potential datasets also provides guidance for third parties to check whether third-party rights are present in a dataset.

Finally, Open Addresses also provides a process through which data holders can report rights infringements. As the provenance of data is recorded throughout the system, this can enable large numbers of addresses to be removed immediately. However this is necessarily a best-effort process: addresses submitted to Open Addresses may appear within the downloadable dataset, and therefore be untraceably copied between publication and redaction.

ODI

open addresses

white paper

Authors:
Jeni Tennison
Ellen Broad

Editor:
Anna Scott

Production:
Phil Lang

Design:
Adrian Philpott