

Registers and collaboration: making lists we can trust

Contents

What are registers and why are they important?	4
An introduction to registers	5
The Register Society's Register of Shipping	7
How the internet has changed registers	8
Registers as web services	9
From web service to data infrastructure	9
Registers and abstractions	10
1. Type of list	11
Primary or secondary source	11
Contents	11
Technology	12
2. Authority	13
Sources of authority	13
Reality	15
Completeness	16
Accuracy	16
3. Stewarding	17
Who will maintain this register?	18
How is an entry added to the register?	18
How are errors in the register corrected?	18
Who is paying to maintain this register?	18
Communicating stewarding	18
Collaboration makes better registers	20
Funding registers	21
State funding	21
Service revenue	21
HM Land Registry	22
Grants and donations	23
Building trust when stewarding	24
Models for collaboration	25
Custodial discretion	25
Suggested edits	26
Food Standards Agency Sampling Classification Register	26
Direct access	27
Registering a company in England and Wales	28
Aggregator registers	29
UK Postal Addresses	30
Crowdsourcing	31
Democracy Club Candidates	32
org-id.guide	33
Mixed models	34
Appendix 1: report methodology	34

About

This report is about registers or lists, and ways of publishing and maintaining them. We examine models of stewarding registers and review case studies of different registers. We argue that transparency is necessary, and collaboration desirable, in managing registers that are trusted by their users.

This report has been researched and produced by the Open Data Institute, and published in May 2018. Its lead authors were David Miller and Sym Roe. If you want to share feedback by email or would like to get in touch, contact the Registers project lead Leigh Dodds at Leigh.Dodds@theodi.org.



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

What are registers and why are they important?

Registers are accurate, complete lists. However, not all lists are registers. This document explores some of the key properties of registers and discusses the implications of these. It does not provide a hard definition or set of criteria for what is or is not a register. Such exercises have tended to produce criteria that exclude lists that it is helpful for maintainers and users to consider as registers, or the criteria have been so vague that they cease to be meaningful.

What are the key properties of registers beyond accuracy and completeness?

Registers are only useful when we trust them. This means that we need a reason to trust the process of keeping the list up to date, ideally because that process is clear and transparent.

What do registers contain? There are no hard rules about what they contain, but broadly speaking, the contents of registers are important enough for many people to care about them being accurate and complete. For us to think of a list as a register, an institution must have promised to keep it up to date.

Digital registers reduce the cost of services that require data to operate and reduce the burden on institutions that have to manage data. Therefore, publishing registers that take advantage of digital technology can unlock space for new products and services that have previously been too expensive or too risky to build and maintain.

Almost all registers use some degree of collaboration. Embracing that collaboration often increases their quality and usefulness. Today, there are tools that can make that collaborative process faster, cheaper and more accessible. Taking advantage of them is an opportunity to strengthen our data infrastructure¹, at reduced cost.

We define the process of maintaining a register through collaboration as stewarding. This is the process of maintaining a list by a group or institution and the promise they make to maintain that list. That process is co-ordinated by a custodian. Transparent stewarding of a list means being clear and open about how it is maintained.

¹ Open Data Institute, 2018, 'Data Infrastructure', <https://theodi.org/topic/data-infrastructure>

An introduction to registers

The word register is both a noun and a verb. The word register also implies that there is an official quality to the list or act of listing.

For many of us, our first introduction to registers comes as we sit in a classroom and the teacher takes the register. The teacher reads out a list of names and we sit waiting for our own. When we hear it, we answer “yes”. The teacher makes a mark against our name and our attendance is recorded. This simple example actually contains several of the key features of registers – that they are recorded, accurate and official – as well as some less-obvious characteristics.

There are three uses of ‘register’ – two nouns and a verb – in the act of taking the register. Our teacher reads from the register of pupils in this class (1) in order to create the register of attendance (2). When the pupils answer in the affirmative, the teacher will register them (3) in the register of attendance (2). These different uses of ‘register’ will be used throughout this report, though we will avoid using register as a verb except where it has a specific legal meaning. Instead, we talk about creating, updating and removing entries from a register.

The classroom example demonstrates many of the characteristics of registers we discuss in this report. The primary register of the pupils in a class is a subset of the secondary register of pupils at a school. Both of these registers are issued by the school, an institution with authority over the contents of the lists. We know the register of pupils in class 1B is a complete register that contains all the pupils in the class. These registers define reality. When your name appears on the school’s register of pupils, you have become a pupil at this school. Similarly, when a pupil’s name is added to the register of class 1B, they are in class 1B.

The daily register of attendance describes reality. It is a record of who is present that day (and who is not), but it does not change who is in the room or who is a pupil in the class. The register of attendance is created by the teacher, who has the specifically assigned responsibility to undertake the process of researching reality and recording the results of their understanding. This register is also verifiable by anyone else in the room: they too could mark off the names of the pupils and create the register independently. The school will aggregate these individual registers of attendance, creating a list of all the pupils who were at the school on any given day.

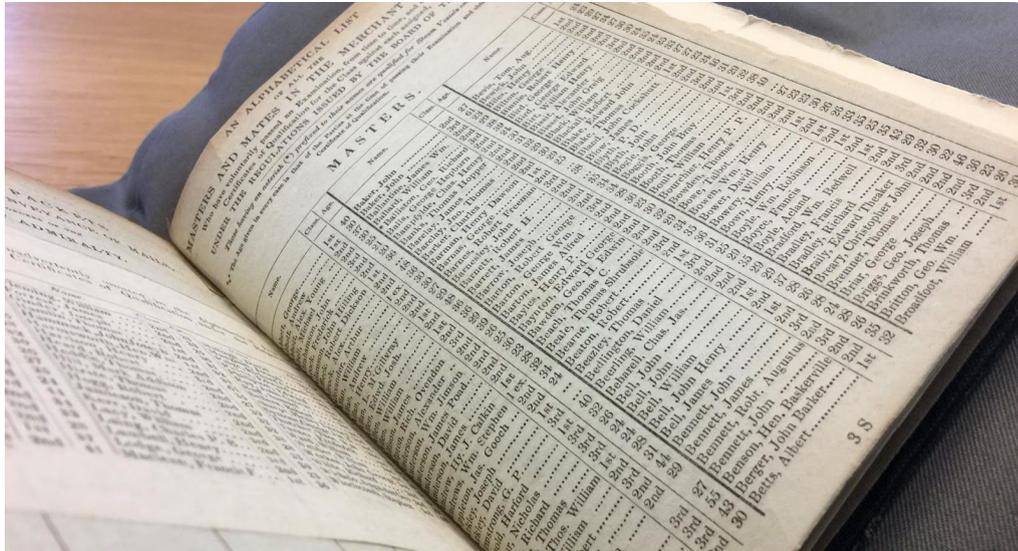
We create registers in order to do something with the information they contain. When a pupil arrives on the first day of school, members of staff will check for their name on the register of pupils in a class to find out which room they should go to. Later on, the register of pupils at the school will be used to find out about overall attendance. This can be divided into individual pupils (who can be given additional help if their attendance is low) or represent the school overall (which contributes to its quality assessment by Ofsted, which in turn is used by parents to work out where to send their children).

Registers have a long history outside the classroom. Some of these registers are very different kinds of lists, particularly with regard to how hard it is to maintain their objective completeness and accuracy.

Case study: The Register Society's Register of Shipping

In the late 18th century, The Register Society created a register of shipping that “classified and registered vessels according to certain criteria of physical structure and equipment, to enable underwriters, shipbrokers, and shipowners more easily to assess commercial risk and to negotiate marine insurance rates”.²

Becoming Lloyd's Register of Shipping in 1834, this register has been published annually since at least 1775. For most of this time, the register was published as a paper book.



The Register of Shipping differs from school registers in several ways. While a register of pupils in a class contains very little information – just the name of the pupil – the Register of Shipping records many different attributes about each ship. Although some of this information is reproducible and verifiable, such as tonnage and dimensions, and can be obtained by surveying, some items in the register are less objectively defined. Over time, the characteristics of ships that people considered important to record has changed: ‘guns’ are no longer a standard ship classification.

Much of the Register of Shipping is a list of opinions that underwriters and owners of ships had about the quality of ships. Because entries in the register were not always things that could be objectively agreed, a dispute between the two groups caused two registers to be published.³ Called The Green Book and The Red Book, each contained different lists of ships and different information about those ships.

The design of the Register of Shipping meant that it could never be completely accurate. It would be legitimate for two people to hold different opinions about a single value in the register. Even with these potential discrepancies, it proved immensely useful to the shipping industry. This official list, which different parties agreed to use as the standard, allowed owners to insure their ships.

² Michael Palmer (1999), ‘Lloyds Register of Shipping’, <http://www.mariners-l.co.uk/ResLloydsRegister.htm>

³ University of Glasgow Archive Services (2011). ‘Lloyd’s Register of Shipping’, https://www.gla.ac.uk/media/media_67062_en.pdf

How the internet has changed registers

The internet and the web introduced new possibilities for registers.

Publishing and updating became easier and quicker. But publishing formats and processes have been slow to adapt to new opportunities and needs. Progress from registers on private databases or (even) paper to registers as part of trusted data infrastructure, available to help other web services deliver reliable online services, is still uneven.

Storing registers in a database allows us to both store lists and to perform complex operations like searching, aggregating and filtering in fractions of a second. Many registers were moved to databases decades before the invention of the web.

When the web became widely available registers started to be published online. They tended to be published as static 'snapshots', often in parallel with printed 'editions' and in formats such as PDF which can be read easily by humans but not by machines.

A bigger step was needed to make registers part of the web, key components of a data infrastructure which enables a host of new, simpler, cheaper and faster services for everything from renewing a passport to global shipping insurance brokerage.

Registers as web services

Registers have started to become web services rather than static collections of data. The register itself, rather than a frozen-in-time snapshot of it, is available online. It is always up to date. And access to it uses the natural patterns and protocols of the internet: register entries have their own addresses on the web and can be manipulated, used or linked to by third parties.

Turning a register into a usable web service requires substantial change by its custodians, affecting its publishing methods and systems and, often, the design and structure of the register itself.

From web service to data infrastructure

Treating registers as infrastructure requires that we can reliably expect and inspect specific standards of authority and continuity. This allows the creation of tools and services built on the data. It makes it easier to curate, maintain and consume data.

Infrastructure, particularly when operated by the state, implies a long-term commitment. Users can assume that there will continue to be an accurate list and that curating it will continue to be someone's specific responsibility.

Reliable data of a known provenance is critical to people who build services using the data held in registers. Being able to rely on a register as infrastructure for their services means they only ever need to work with one reliable and current source of data – what is known in information systems theory as a 'single source of truth'. They can concentrate on building the service rather than spending time checking if the data they are using is still accurate.

When registers can be easily accessed by other products or services, they are more widely useful than when they were books in council basements. Combining internet technology for publication with a commitment from the state to maintain registers means society can reap increased benefits from these official lists of things, consuming and using them in ways that were not previously possible.

Registers and abstractions

Before exploring the details of specific registers, it is worth looking at some of the abstract aspects of lists. We are deliberately examining the more general case of lists rather than only registers. This helps establish ways of talking and thinking about different kinds of registers as well as allowing us to consider which of these properties are required for us to think of a list as a register.

We have grouped the features of registers into three related categories: type of list, authority and stewardship. We created the categories by examining a sample of 20 registers of varying forms, how they were similar, how they differed and by noting how people discuss them and describe their key features. The results of this analysis are available as open data.⁴ When people make lists, they often create hybrid forms that cross these boundaries.

1. Type of list

The first group of features allows us to understand what kind of list we are dealing with.

Primary or secondary source

Some lists will be the primary source of that information; others will not. These are usually an aggregation of smaller lists. The register of attendance taken by a teacher is the primary source of data about which pupils attended class that day. The school-wide attendance register is an aggregated, secondary list made up of all the other (smaller) lists of who attended class.

In this report, we refer to source lists as primary and aggregated/derived ones as secondary.

Primary: the main source of the information contained in it, like a class attendance register, or the register of companies provided by Companies House.⁵

Secondary: A list created from other sources of information, like the school-wide attendance register, or the database of companies provided by Open Corporates.⁶

Contents: The level of detail that we include about each entry in a list, e.g. its name, identifier and other attributes, is a central aspect of how we think and talk about the list. When designing a list, we have to decide how much information to record for

⁴ David Miller, Sym Roe (2018), 'Registers and collaboration - describing lists', <https://docs.google.com/spreadsheets/d/1eLmRzkKFJx6hpOcVTv6WmcFYOULCB0SWSOCN0IF5O1Q/edit>

⁵ Companies House, 'Search the register', <https://beta.companieshouse.gov.uk/>

⁶ 'Open Corporates', <http://opencorporates.com/>. Open Corporates reuse data published by Companies House and other sources.

each entry. This decision sits on a sliding scale, ranging from least possible amount of information to most possible amount of information.

This scale is highly subjective. Trying to define hard criteria or subsections of the scale is not particularly useful. Instead, think about lists in terms of which end of the scale they are closest to, from minimum viable dataset to what we call a broad dataset.

Minimum viable dataset: tightly defined, minimal schema of what makes up a dataset. There is an expectation that all or at least most attributes will be completed for each entry. The FCO Country register⁷ or the register of pupils are closer to this end of the scale.

Broad dataset: everything we know or might know about one entry on the list, including attributes that might be based on opinion rather than objective facts. The Shipping Register or Open Street Map⁸ are both close to this end of the scale.

Technology

The technology the list uses to record or publish information can change how we use or think about the list. List technologies fall into the following broad categories:

Paper: the list is stored and published on paper, like the first few hundred years of the Lloyd's Register of Ships.

Digital paper: the list is stored or published electronically, but it uses a format where individual entries are not addressable and the contents not machine-readable. Legal notices published as PDFs or Microsoft Word documents on council websites fall into this broad category.

Machine-readable: the list is published in a machine-readable, computable format, probably online. For example AddressBase, the Ordnance Survey's address data product.⁹

Linkable: the list is published in a machine-readable, computable format, with individual entries addressable via URIs, such as the government's service register.¹⁰

Many lists are published in ways that span more than one of these list technologies; the Companies House register can be downloaded in a machine-readable format¹¹ and is also published as a linkable online service.¹²

⁷ Gov.uk, 'Country register', <https://country.register.gov.uk/>

⁸ OpenStreetmap, <https://www.openstreetmap.org>

⁹ Ordnance Survey, 'AddressBase', <https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase.html>

¹⁰ Gov.uk, 'Government service register', <https://government-service.register.gov.uk/>

¹¹ Companies House, 'Free Companies Data Product', http://download.companieshouse.gov.uk/en_output.html

¹² Companies House, 'Open Data Institute', <https://beta.companieshouse.gov.uk/company/08030289>

List	Source	Contents	Technology
Register of Shipping	Primary	Broad dataset	Paper
Democracy Club Candidates	Secondary	Broad dataset	Linkable
Council candidate notifications	Primary	Minimum viable dataset	Digital paper
Countries Register	Primary	Minimum viable dataset	Linkable
Register of pupils in a class	Primary	Minimum viable dataset	Paper
Historic England's list of listed buildings	Primary	Broad dataset	Machine-readable

Taken from the full list of lists used to make this system¹³.

2. Authority

The second group of features are reasons to consider a list authoritative.

Sources of authority

When talking about complete and accurate lists (or even lists that are complete and accurate enough to use) we also need to be able to understand how we are able to make that assertion. There are various sources of authority:

From first principles: Some lists are provably complete and accurate. A list of the first thousand prime numbers could be proved complete through calculating those numbers and then checking the register against the list you computed.

Verifiable observation: When the list contains things that have been observed in the world, it is often possible for someone else to recreate the list from scratch. Wikipedia's list of rivers longer than 1000 km¹⁴ could be verified by someone else measuring the lengths of rivers.

Observation might introduce bias or opinion in the register in ways that make it a matter of opinion whether or not it can be reproduced. This is a form of observer-expectancy effect.¹⁵ In the classroom register, a teacher might make allowances for a child who is not actually in the room when the register is taken,

¹³ David Miller, Sym Roe (2018), 'Registers and collaboration - describing lists', <https://docs.google.com/spreadsheets/d/1eLmRzkKFJx6hpOcVTv6WmcFYOULCB0SWSOCN0IF5O1Q/edit>

¹⁴ Wikipedia, 'List of rivers longer than 1000 km', https://en.wikipedia.org/wiki/List_of_rivers_by_length#List_of_rivers_longer_than_1000_km

¹⁵ Wikipedia, 'Observer-expectancy effect', https://en.wikipedia.org/wiki/Observer-expectancy_effect

because the pupil is known to be running late. Other teachers might decide to only note pupils who are actually in the room at the exact time when they are taking the register.

Trusted observation: For some lists that are created by observing the world, re-observation or verification is not possible. This is often because they take a snapshot of a point in time. While the school class attendance register can be validated at the time by a third party, a week later it is impossible to verify that it is correct. There might, however, be ways to detect whether there are likely to be errors in a list. In the class attendance register example, if an attendance register on a given day was very different from the same register on surrounding days, you might expect it to be wrong.

The authority of such lists depends on how much we trust both whoever made the observation, and the integrity of the processes for storing, retrieving and publishing the list.

Institutional authority: The contents of the list are the responsibility of a single institution, which asserts that the list is complete and accurate. When Microsoft publishes the list of software it provides support for,¹⁶ we believe this list to be complete. Microsoft can be trusted to say that this is a complete list because it is in a position to unilaterally decide the contents of the list.

Legislative custodian: A subset of institutional responsibility where the institution responsible for maintaining the list is doing so due to a statutory obligation. For instance, the register of UK political parties is maintained by the Electoral Commission¹⁷ due to their obligations in, for example, the Political Parties, Elections and Referendums Act 2000.¹⁸

Trusted custodian: Often, where a list does not fall under any of these categories, its sources of authority are highly subjective, based upon how much you trust the organisation that has put together the list to have done a good job. This makes it difficult to have a truly authoritative list. When looking for a list of companies to use, some people might trust Open Corporates.¹⁹ Others might choose to trust Thomson Reuters²⁰ or Dun and Bradstreet.²¹

None of these categories are mutually exclusive, and there are occasions where a register draws on multiple sources of authority.

Reality

Lists have different relationships to how we understand the world. While the classroom attendance register describes which children were in the room, being on the register of children in class 1B makes a child a member of class 1B.

¹⁶ Microsoft, 'Search product lifecycle',

<https://support.microsoft.com/en-gb/lifecycle/search?ts=-3>

¹⁷ Electoral Commission, 'Where can I find a list of all political parties?',

<http://www.electoralcommission.org.uk/faq/donations-to-political-parties/where-can-i-find-a-list-of-all-political-parties>

¹⁸ Legislation.gov.uk (2018), 'Political Parties, Elections and Referendums Act 2000',

<https://www.legislation.gov.uk/ukpga/2000/41/contents>

¹⁹ Opencorporates, <http://opencorporates.com/>

²⁰ Thomson Reuters (2018), 'Company Data',

<https://financial.thomsonreuters.com/en/products/data-analytics/company-data.html>

²¹ Dun and Bradstreet (2018), 'Our data differentiators',

<https://www.dnb.co.uk/about-us/our-data.html>

We refer to these types of lists as being reality describing or reality defining.

Reality defining: a list that somehow changes reality when an entry appears on it, like the register of pupils in a class or the register of medical practitioners.²² This is generally associated with the source of authority of the list. The source of authority is usually either institutional authority, or the legislation that mandates the existence of the list.

Strictly speaking, these lists do not define reality. They are generally ones that a large number of people have agreed to recognise as authoritative. Therefore, they are also associated with rigorous controls and processes around adding, deleting or changing entries.

Reality describing: a list that describes the world as observed, like a class register of attendance or the Wikidata list of countries in the European Union.²³

There may also be hybrids of these two types, where authoritative, canonical information (reality defining) is supplemented by additional descriptive information (reality describing) in a list.

Completeness

A list may be either complete or incomplete. An incomplete list does not contain all the entities that should be on it, like a list of pupils in a class that is missing some of the children that are meant to be there. A list can also be incomplete because it is missing information about each entity, like a list of contact details which contains a row of names, but is missing the email and phone number of each person.

The register of pupils is complete because it contains all the pupils who are meant to be in the class at that time. By contrast, the Wikipedia list of deaths in 1989²⁴ is incomplete (at the time of writing) because it does not include everyone who died in 1989 and the register of government cats²⁵ is similarly incomplete as it only contains those cats that have been reported.

Accuracy

A list may be entirely accurate, have some known errors, or contain some elements where the provenance and accuracy are unknown.

We would expect a list of numbers between 1 and 100 to be entirely accurate because it is easy to compile and check. A list of every postal address in the country is likely to be inaccurate as new houses are built and other changes occur. Inaccuracy is particularly an issue for reality describing lists.

A list may also be accurate only at a certain granularity. For instance, Open Street Map²⁶ will be less accurate than the Ordnance Survey MasterMap²⁷ product, because

²² General Medical Council (2018), 'Registration and licensing', https://www.gmc-uk.org/doctors/medical_register.asp

²³ Wikipedia (2018), 'Member state of the European Union', https://en.wikipedia.org/wiki/Member_state_of_the_European_Union

²⁴ Wikipedia (2018), '1989', <https://en.wikipedia.org/wiki/1989#Deaths>

²⁵ Peter Wells (2016), 'UK Government cats', <https://peterkwells.github.io/uk-government-cats/>

²⁶ OpenStreetMap, <https://www.openstreetmap.org>

the process of recording information is not performed to the same level of accuracy as that required of the surveyors that Ordnance Survey use.

Some lists contain pieces of information which can never really be entirely accurate; they are opinions rather than facts, as in the Shipping Register example discussed above.

List	Source of authority	Reality	Completeness	Accuracy
Pupils in a class	Institutional	Defining	Complete	Accurate
Open Street Map	Verifiable Observations	Describing	Incomplete	Sometimes accurate
General Medical Council Register	Legislative	Defining	Complete	Accurate
First thousand prime numbers	A priori	Describing	Provably complete	Provably accurate

3. Stewardship

Stewarding is the process of maintaining a register. This includes how and when information is added, removed or changed and who does those things. Understanding how a custodian stewards a register allows people who are thinking about using it to make informed decisions about how much they can trust it.

We explore the stewardship of a register by asking a series of questions.

Who will maintain this register?

Which individual, group or institution is responsible for making sure it remains accurate and complete?

How is an entry added to the register?

What is the process for adding an entry to the list? Who is involved? Will there be an approval step or can something be added directly?

How are errors in the register corrected?

Registers often contain errors, omissions or inaccuracies. This is particularly true of registers which are reality describing, as they are susceptible to external changes which are not yet reflected in the register. When this happens, there needs to be a way to make changes to that register.

²⁷ Ordnance Survey (2018), 'OS MasterMap', <https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html>

Who is paying to maintain this register?

Making sure a register remains accurate and complete over time requires work. This often requires funding to pay for staff or infrastructure.

Understanding who pays for such work is important. Funders often maintain a register for a particular purpose. Considering what these purposes are can help us understand how the register is governed.

Communicating stewarding

The benefits of register infrastructure come from the contents of the register being used by many people, systems or institutions. To achieve this, the register needs to be seen as trustworthy and authoritative by those users.

If users cannot see how the register is governed, compiled and quality-assured, they cannot know if it is appropriate to reuse. Such assessments include whether it will be accurate enough, whether the precise and detailed definition of what constitutes an entry on the register is suitable for a specific purpose, as well as other questions based on the details of the particular subject matter.

Most registers and lists that claim to be authoritative and suitable for reuse do not explain the process of stewarding, or only partially explain it. While it is often possible to find out how the stewarding works, doing so can often require extensive research. This increases the cost of reuse, which in turn probably reduces how often it is reused.

For example, at the time of writing the Foreign and Commonwealth Office's register of countries recognised by the UK²⁸ does not give a clear picture of its stewarding.

The register website simply lists the government department responsible for maintaining it. This leaves many questions unanswered. The register provides no way of contacting the department to report errors or suggest changes to the register. It has no information about how decisions have been made about which countries are included. The register is not simply "British English-language names and descriptive terms for countries" (the description given in the register itself); it contains countries formally recognised by the government of the United Kingdom.

²⁸ Gov.uk, 'Country register', <https://country.register.gov.uk>

Country register

British English-language names and descriptive terms for countries

[View Register](#)

Managed by



Foreign &
Commonwealth
Office

Find out more

[About registers](#)

Using this register

Base URL

`https://country.register.gov.uk`

Endpoints

All endpoints used to interact with a register are available as [JSON](#), [YAML](#), [TTL](#), [CSV](#) and [TSV](#).

You can find the correct path for the query that interests you in our [API Reference](#). Users should append the path to the base URL and specify a preferred format, such as [JSON](#).

`/records` returns all records, the latest version of the data in the register. A record is the most recent entry for a key and its associated items.

Without understanding how and why a country has an entry on the register, it is hard for anyone considering using the register to decide whether or not it is an appropriate list of countries. This in turn influences what you can use the register for. A service in the UK recognising passports or visas could not use this country register as it does not include Taiwan (the UK does not formally recognise Taiwan). However, Taiwan issues passports to citizens, and they are granted visas to travel to the UK.

Obscure stewarding is currently a feature of all government registers; there is no publicly available information about how such lists are maintained.

Collaboration makes better registers

Registers are established for a purpose, but it is rare for even the main intended users to be involved in the specific design of the register – let alone people who may not be the intended users but who have a legitimate use for the information it contains.

In government, failing to consult or co-design with users to make the register as easy and relevant as possible adds to the costs of the nation and reduces its efficiency.

In commerce, the design of a register should follow the processes and procedures recommended for other kinds of open standards for data, and for the same reasons²⁹.

²⁹ Open Data Institute (2018), 'Exploring the development and impact of open standards for data report', <https://docs.google.com/document/d/1Sab5YMVj4PVqLjZD35hX8FTnMeeP6gLGG0xszuRMIaM/edit#heading=h.zenpzz1aus85>

This call for transparency and collaboration does not mean that registers should accept every suggestion made, or that they should attempt to meet the needs of everyone.

Meeting the needs of data users is not the same as meeting the needs of the users of the services built using data.

For registers designed as data infrastructure (due to the social, legal, environmental or economic value of their contents), explaining their governance processes is a minimum requirement.

Those processes should be both transparent and open. The criteria by which decisions are made, and also individual decisions, should be public. Where such decisions are potentially controversial, or subjective, someone who disagrees with the contents of a register should at least be able to understand why the maintainer of a register has made their decision.

Funding registers

Understanding who is paying for something can give valuable insights into the governance of the register. It also may help us to make judgements about things like

- is it likely to be sustained in the long term?
- is it likely to be unbiased?
- are its operations likely to be well-funded?

Traditionally, registers were often maintained by an organ of the state (although there are exceptions, such as Lloyd's Register). This is related to practical aspects of running a state, such as taxation, and to the fact that the creating and maintaining registers was both time-consuming and expensive.

Today, there are three main ways that registers are funded:

- State funding
- Service revenue
- Grants and donations

Registers are often funded through a mixture of these.

State funding

Registers operated by the state are generally funded by taxation. These registers may be directly operated by a government body or by a third party the state enlists to maintain the register.

Service revenue

Sometimes, institutions who are paid to maintain a register are also instructed to generate revenues from the register. This supports the maintenance of the register itself and sometimes provides additional revenue for the Treasury. There are three typical models: charging for updates to the register, selling commercial services related to that data, or charging for access to the data within the register. These three models have differing levels of tension between providing that data for free at the point of use and revenue maximisation.

Case study: HM Land Registry

HM Land Registry was established by the Land Registry Act 1862.³⁰ Since then, it has registered ownership of land and property in England and Wales. Registration of land was initially voluntary, though compulsory registration was gradually introduced and extended over the next 140 years.

The registry took from 1862 to 1963 to collect two million entries.³¹ Since then, the process has accelerated. The Land Registry now contains more than 24 million titles. This acceleration has been driven by the increasing speed of processing introduced by digitisation of the registry.

Registration is only compulsory under certain circumstances, such as when property is sold or inherited. Therefore, the registry still does not cover the whole of England and Wales. As of 2017, HM Land Registry claimed to have 85% coverage.³²

While HM Land Registry is a department of the government of the United Kingdom, they do not receive public funding. Rather, they generate revenue by charging people who want to change or query the register. In 2016, this revenue was more than £300M. This allows HM Land Registry to operate in the way that it does: backed by legislative authority without needing to be funded by taxes.

³⁰ Legislation.gov.uk, 'Land Registry Act 1862',
<http://www.legislation.gov.uk/ukpga/Vict/25-26/53>

³¹ Land Registry (2000), 'A Short History of Land Registration in England and Wales',
<http://webarchive.nationalarchives.gov.uk/20101213175739/http://www.landreg.gov.uk/assets/library/documents/bhist-lr.pdf>

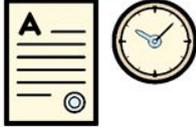
³² Land Registry (2018), 'About Us',
<https://www.gov.uk/government/organisations/land-registry/about>



Register change services

Update existing title

£155.5m



Intake volume
5.0m

Leasehold creation from existing title

£30.6m



Intake volume
0.2m

Splitting freehold titles

£32.1m



Intake volume
0.2m

New title creation

£8.3m



Intake volume
0.1m



Register query services

Views of the register



£19.1m

Intake volume
6.4m

MapSearch (PDF)



Intake volume
1.1m

Official copies of the register



£44.0m

Intake volume
14.6m

Guaranteed search of the index map



£1.1m

Intake volume
0.8m

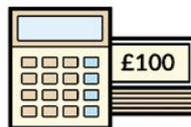
Official searches of the register



£2.4m

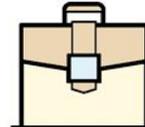
Intake volume
2.4m

▶ Other services



Land Charges

£6.3m



Commercial services

£5.6m

5

Grants and donations

As the internet has made new modes of collaboration and coordination possible, increasingly there are institutions capable of creating reasonably authoritative, complete registers without requiring state funding and complex organisational overheads. Non-governmental organisations³³ (NGOs) and charities³⁴ are seeking to effect change by maintaining digital registers that are in the public interest. That said, keeping such registers accurate and complete still requires some organisation and, often, funding. They are often funded in the same way as other work that charities and NGOs perform, such as grant funding, donations and through volunteers donating their time and skills.

³³ Eg Democracy Club, 'Democracy Club Candidates', <https://candidates.democracyclub.org.uk/>

³⁴ Eg MySociety, 'TheyWorkForYou All MPs', <https://www.theyworkforyou.com/mps/>

We even see instances where complete and accurate lists of things can be maintained by groups of volunteers without paid-for central professional stewarding. A simple example is the list of US State birds on Wikipedia.³⁵ This is particularly true of lists which can be verified by observation and comprise a tightly constrained minimum viable dataset. For some such registers (often small to moderate in size) the overheads relating to publishing them online, such as server space, bandwidth and similar costs, are close to zero.

³⁵ Wikipedia (2018), 'List of U.S. state birds', https://en.wikipedia.org/wiki/List_of_U.S._state_birds

Building trust when stewarding

Trust in registers is built by high-quality stewarding practices. While much of the activity of a custodian comprises ensuring that the register is kept up to date and accurate, it is also the responsibility of the custodian to make the governance processes surrounding a register transparent and open.

For many registers, improving the transparency that surrounds these processes is the simplest and cheapest way to increase trust in their contents.

Transparent stewarding means being clear and open about the answers to the four key governance questions outlined above:

- Who maintains this register and why?
- How are entries added to the register?
- How are errors in the register corrected?
- Who pays for the maintenance of this register?

Models for collaboration

We have observed a number of different models for collaboration to make and maintain registers. We have examined some of these models as case studies to show how they work.

Custodial discretion

For some registers, there is an obvious decision maker who unilaterally operates upon the contents of the register. They most commonly have either legislative custodians or institutional authority.

When Google decides to stop supporting the Mosaic web browser and removes it from their list of supported web browsers, this is a decision taken by a person or group that has been given executive power. This results in Mosaic being removed from the list of supported web browsers.

If you are not the custodian, your ability to collaborate on maintaining the register is dramatically reduced. There may be a clear mechanism to provide feedback, which provides some opportunity to request additions or corrections, but this is not always the case. Registers that operate a custodial discretion model may just as easily operate on a whim, or make decisions in a smoke-filled room with minimal transparency of the precise process.

While errors can be fixed or changes made by the right person in the organisation that created the list, finding that person is often difficult. Formal error-reporting mechanisms may not exist. This model is poorly suited to reality-describing registers, which often need to change rapidly, as the custodian acts as a bottleneck for changes. This in turn makes it hard for the register to be kept up to date; it also increases the resources required because they cannot take advantage of other people who might be able to observe reality. Depending on the level of transparency, a lack of timeliness and an unclear process can diminish trust in the register itself.

Custodial discretion can be effective for reality-altering registers when the custodian is always involved in the change process.

Suggested edits

In other cases, the custodian will make suggested alterations to the register when another party requests them, but the custodian verifies the accuracy of the alteration before making a change. This allows a broad collaboration with many parties on the contents of the register, while maintaining a narrow responsibility for accuracy.

One of the benefits of publishing registers in an online, linkable way is that it makes it far easier to examine their contents for inaccuracies. When tightly coupled with error reporting or change suggestion tools, this way of publishing registers unlocks

potential for collaboration and error detection at a rate that was unavailable to registers either on paper or held electronically without being linkable.

Food Standards Agency Sampling Classification Register

The Food Standards Agency (FSA) is responsible for the Food Surveillance System.³⁶ This is a national database of analytical results for food samples as part of official controls. Sampling is conducted by enforcement authorities, local authorities and port health authorities.

The sample programme requires that samples are rigorously classified in a consistent manner across the country. This classification was previously managed as part of the data capture software product. Any change to the classification involved a complex gathering of requirements and a change management process. Once this was completed, and a new release of the data capture software made, each of the individual laboratories testing samples on behalf of local authorities would have to manage and implement the relevant software update – an expensive and time-consuming process.

Since becoming a register on the FSA registry³⁷ this food sampling classification has become easier. The online register provides a place where any lab that conducts sampling, or indeed any party interested in food sampling, can download the full list. The sampling classification has become a data product, instead of being a small feature hidden in a broader product with unrelated change processes, priorities and incentives.

This shift has enabled the FSA to make engaging with laboratories to improve the sampling classification register much easier. They have taken direct ownership of the change processes by which food sampling is classified, making it possible to change things much more dynamically.

The publication technology makes the complete register available online. It also provides a path for people to submit and track changes. As a result, the FSA has been able to gather improvements to the information from experts in the field, without having to pay someone to travel across the country and speak to them in person.

By abstracting the data standards required in the surveillance system as their own register, the FSA has decoupled data collection from classification. This gives sampling laboratories a choice of how to collect and submit this data. They can innovate with their own data collection and audit processes without being tied to a single monopoly supplier.

Direct access

The most permissive change model is providing tools that allow anyone to directly alter the register. These tools may be an online service (like a wiki), or paper forms that anyone can fill in and send to the custodian responsible for compiling or publishing the register.

³⁶ Food Standards Agency (2018), 'UK Food Surveillance System',

<https://www.food.gov.uk/enforcement/sampling/fss>

³⁷ Food Standards Agency (2018), 'Register: Classifications',

<https://data.food.gov.uk/codes/enforcement-monitoring/sampling/classifications>

With a direct access model, there is no validation of the information provided. There may, however, be validation that the information is within certain parameters. There might not be any examination of the dates on birth certificates or passports, as long as these are dates that could be correct (for example, not in the future).

Direct access models have to provide some mechanism to deal with vandalism or fraudulent entries. They frequently include a proactive mechanism for reporting inaccuracies. Especially with official registers that carry some legal weight, there may be penalties for deliberately entering misinformation.

Where registers using this model have been created in the recent past, they have tended to be created and evolved alongside tools explicitly designed to facilitate these governance processes. This frequently includes some sort of version control system that allows other users (sometimes only with special permissions) to roll back malicious or accidental changes. The penalty for vandalism in such cases is often a ban on a particular user.

In such cases, the governance model and the register are so closely linked that they are almost the same thing; Wikidata as a product is both an editing platform and a data platform.

Registering a company in England and Wales

Companies House is the public body responsible for incorporating and dissolving companies. They register company information and make it available to the public.³⁸ The Companies House register of companies is a reality-altering register with legislative authority.

If a person wants to start a company, they are required to register that company. To do so they must provide some information. For a private limited company, you have to submit the company name, address, directors and details of shareholders to Companies House, using either an online service³⁹ or by post with a paper form.⁴⁰ Some of these details must comply with certain rules.⁴¹ For example, you cannot choose a company name that implies a connection with government or local authorities without permission.

However, Companies House do not perform any additional checks on the accuracy of the information that is being submitted. If a mistake is made on a form, Companies House will add the incorrect information to the register. At times this happens even when the information is known to be inaccurate at the time of submission: “Where a form is completed incorrectly we are not able to reject the form”.⁴²

As it is an official register, the company directors have a duty to correct any errors or inconsistencies they are aware of. In some cases, the registrar may mark a record as “inconsistent”. In extreme cases, where no correction has been made, these can constitute an offence and may be liable for a fine.

³⁸ Companies House (2018), ‘About Us’, <https://www.gov.uk/government/organisations/companies-house/about>

³⁹ Companies House (2018), ‘Web Filing’, <https://ewf.companieshouse.gov.uk//seclogin?tc=1>

⁴⁰ Gov.uk (2018), ‘Register a private or public company (IN01)’, <https://www.gov.uk/government/publications/register-a-private-or-public-company-in01>

⁴¹ Gov.uk (2018), ‘Set up a private limited company’, <https://www.gov.uk/limited-company-formation/choose-company-name>

⁴² Companies House (2015), ‘Inconsistencies on the public register’, <https://www.gov.uk/government/publications/inconsistencies-on-the-public-register/inconsistencies-on-the-public-register>

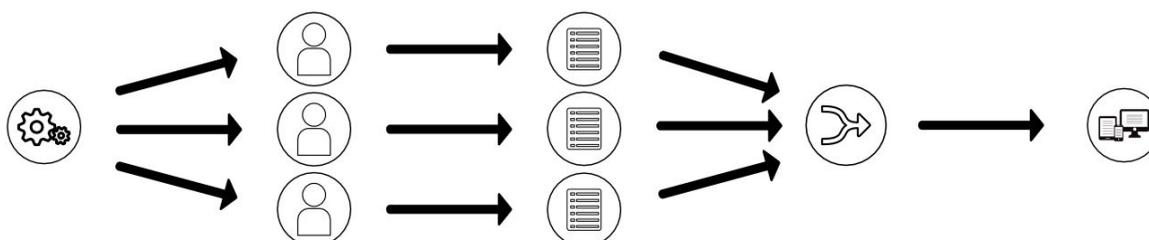
Companies House receives no taxpayer funds. It is mostly funded by charging fees for company registrations. A smaller proportion of its income comes from charging people who want to receive information from the register on paper or in bulk, and by renting buildings they own to other parts of government.⁴³

Aggregator registers

A secondary register is a list created from other registers, for instance by aggregating other smaller registers.

In this model, many different custodians create separate registers, which are then be merged into a single authoritative register. This often occurs where, due to either historical accident or design, there is strong institutional authority or legislative custodian of lists of data that are a subset of a larger group of identical or very similar things.

Separate custodians are also found where there is a division of the register that makes logical or administrative sense. In this manner, book publishers are allocated blocks of ISBN numbers⁴⁴ to use for books they publish.



Due to the range of ways in which current custodians make registers they are responsible for available, automating the transformation of primary registers into a secondary register is not always possible.

UK Postal Addresses

The UK does not have a single authoritative list of geographic postal addresses. It does have a de facto ‘most-trusted’ list called AddressBase, which is one of a range of commercial products sold by Ordnance Survey, a government agency. The process of creating the address list is complex. There are many sources, each with slightly different reasons for contributing to it. As a result, it is difficult to talk about addresses as a single list. It is possible that it would be better understood as a set of smaller lists that could be combined depending on use.

Each organisation’s section of the list comes with slightly different properties. For example, the Royal Mail contributes postcodes. These are designed to help someone

⁴³ Companies House (2016), ‘Annual Report and Accounts 2015/16’, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/540443/AnnualReport_201516.pdf

⁴⁴ Wikipedia (2018), ‘International Standard Book Number’, https://en.wikipedia.org/wiki/International_Standard_Book_Number#Registrant_element

deliver letters to groups of buildings. Carrying a letter to a house has different requirements than knowing the exact latitude and longitude of each property, or knowing whether the property is a hostel with multiple residents, or the exact name of each street.

Each local authority contributes to the list via the Local Land and Property Gazetteer,⁴⁵ which is in turn made by street naming and house numbering officers, and which also has other planning and tax functions.

Most other elements of the list are reality describing. There are some exceptions like postcodes, which are assigned to groups of houses, thus altering reality. New addresses at the local authority level are a form of reality-altering register as they define what the street names and house numbers will be. They quickly become reality-describing registers as people rename houses, split them into flats and so on.

All these sublists, with each of their different governance and custodial patterns, are aggregated by GeoPlace LLP. They are then turned into AddressBase, sold by Ordnance Survey (OS), and PAF, sold by Royal Mail.

OS charges for access to AddressBase and has a form of error reporting governance model. However, because it is a secondary register, error reports might target the wrong institution, leaving Ordnance Survey or GeoPlace to relay error reports on to upstream data maintainers.

The list technology is machine-readable but not linkable. The data is only published as snapshots, roughly every six weeks.

As a reality-describing, secondary register that is very large with a lot of updates, AddressBase is prone to a lot of errors and will always be slightly out of date. It is possible that no list of this sort could ever be considered “complete”. Because of this, easy and transparent error reporting is vital.

Crowdsourcing

Crowdsourcing is a model of sourcing information from a community. Typically, the total information required has been broken down into small chunks. These can be sourced with relatively little effort and then the results are aggregated. This model typically employs a strategy to prevent abuse or vandalism. One such strategy is post-moderation or version control combined with tooling like the ability to monitor content changes or being notified of edits, allowing the community to maintain quality. Another is to require multiple people to provide the same information – providing a form of verified observation.

An editable Google Document is a simple tool for crowdsourcing, with fewer features for community maintenance, while Wikipedia and OpenStreetMap are example of large scale, moderation-based crowdsourcing platforms.

Democracy Club Candidates

Democracy Club Candidates⁴⁶ is a database of candidates for elected office in the UK.

⁴⁵ Wikipedia (2018), ‘Local Land and Property Gazetteer’, https://en.wikipedia.org/wiki/Local_Land_and_Property_Gazetteer

⁴⁶ DemocracyClub (2018), ‘Candidates’, <http://candidates.democracyclub.org.uk/>

For elections covered in the past, the list of candidate names and parties are verifiably complete because each local authority publishes this data as an authoritative, reality-describing, non-machine-readable, hard-to-discover list called a Statement Of Persons Nominated (SOPN).

Democracy Club provides a data service on top of these siloed lists through both combining them and making them machine-readable. Adding a person to Democracy Club's list does not mean they are nominated, but being on a SOPN does.

For future elections Democracy Club maintains a sourced, non-canonical, patchy list of people they think will end up on the verifiably complete list. The number of attributes in this speculative list is greater than the official registers (for example, it includes Twitter handles).

Each field is crowdsourced and each edit requires a source. This means the database is reproducible by anyone willing to invest the same effort. Particularly for information that is not sourced from official publications or formal notices from local authorities, observations may not be verifiable after the fact. For instance, candidates often take down websites they operated for their campaign.

This means that the list of all candidates Democracy Club knows about contains many different types of list with content that varies from verifiably complete, to rough assertions.

Some fields in the list use different patterns for contributions depending on the type of field and its importance when converted from a sourced, non-canonical, patchy list to a verifiably complete list. Photos are pre-moderated to avoid malicious, hard-to-automatically-select uploads. Once a candidate has been verified against a SOPN – by at least two people checking – you cannot edit the candidate's name and party. This “locking” is represented in the API and CSV downloads, which allows data consumers to decide on the type of list they want.

org-id.guide

Org-id.guide⁴⁷ is a register where each entry is itself a list of organisations. It allows people to locate unique identifiers for organisations around the world by providing a canonical form to refer to the many different sources of unique identifier for organisations that we find in the world.

It does not store unique identifiers itself as an aggregator. Instead, it provides an index of external primary registers where such things may be found. It also provides additional metadata to enable the use of open, interoperable, unambiguous organisation identifiers.

Org-id.guide provides a contributor's handbook⁴⁸ for people looking to suggest an addition or change.

This is an official way for people to either request or propose changes. The process is managed using the Github platform. Requests are raised as Github issues, and proposals as Github pull requests. There are clear guidelines for the details a request or proposal should include. The results of decisions made in reaction to requests or proposals are made in the open, and can be reviewed by anyone. Each decision has

⁴⁷ Open Data Services Co-op, 'Org-id.guide', <http://org-id.guide/>

⁴⁸ Open Data Services Co-op, 'Org-id.guide contributors handbook', <http://docs.org-id.guide/en/latest/contribute/>

a URI. This also contains information about the discussions that contributed to any given decision, and the names of the people who eventually accepted or rejected a suggested alteration. The decision-making process is also outlined, clarifying whose responsibility it is to decide on a proposal, how to provide feedback on somebody else's proposal and the amount of time available to provide such feedback.

Mixed models

In practice, many registers operate a mixed governance model. This allows maintainers to spend their time more effectively. They can focus on particularly complex or sensitive tasks or areas, while allowing other people to effectively collaborate on maintaining the register.

Companies House allows essentially direct access to many operations, such as creating a new company or recording a share issue. A form is submitted, but not checked beyond minimal validation. However, when a company is struck off the register, only the custodian can record that piece of information.

Similarly, with crowdsourced lists such as Wikidata, the default is direct access to create, alter or remove entries, but certain pages are protected.⁴⁹ This changes the default behaviour, as only administrators can make changes to those pages.

⁴⁹ Wikidata (2018), 'Wikidata:Page protection policy', https://www.wikidata.org/wiki/Wikidata:Page_protection_policy

Appendix 1: report methodology

This report has been written by David Miller and Sym Roe in partnership with the ODI.

It draws on their experience of open data and registers, desk research into current and historical practices around registers, and interviews with people who self-identify as being interested in registers and open data.

The interviews were semi-structured, with questions designed to stimulate conversation. Interviewees were asked questions such as “can you define a register?”, “can you give an example of a list that isn’t a register?” and “what would you change if you had a magic wand?”

Interviewees were recruited through David, Sym and the ODI’s existing network of experts. Interviews included representatives from Epimorphics, Food Standards Agency, Government Digital Service, Office of National Statistics, Ordnance Survey, Open Data Services, Thomson Reuters and Wikidata.

Time constraints meant that no wider recruitment or quantitative research was done. This report should be considered a starting point for debate rather than evidence of any point or argument.

All internet links in footnotes were active at the time of publication. Should the reader find that links are no longer active we would recommend consulting the Internet Archive.⁵⁰

⁵⁰ Internet Archive, <https://archive.org>