Achieving more with open data at data.parliament.uk

Understanding the user need for Parliamentary data

Authors: Thomas Forth (ODI Leeds) and Richard Norris (the ODI)

Summary

In the course of our research we have found that people who work with data from Parliament trust it to be objective and definitive. They consider Parliament itself to be uniquely placed to report and curate what happens there, and want it to **focus on providing better data ahead of building more tools**.

Parliament is in the best position to create data models that represent its functions accurately and objectively, populate these models, and then make this available as open data. Adding Open Data Certificates to these datasets will provide a mark of quality and trust to users.

A considerable amount of work is already being done within data.parliament.uk to provide what users are asking for. It is essential that Parliament continues to listen to feedback and **deliver short-term improvements**, whilst also building and sharing a comprehensive data model for Parliament.

While we urge data.parliament.uk to focus on providing excellent data, the positive responses to the new Hansard and parliamentlive.tv sites show that tools matter too. **Building tools and websites on Parliament open data is a good way to test and improve the quality of the open data**.

The majority of public interactions with Parliament are indirect, via tools like TheyWorkForYou and via journalists who often rely on these secondary sources.

We recommend that Parliament **supports the network of data users and developers** who create tools that extend its reach. Understanding the needs of end users is also important, and Parliament should work with its network to do this.

Open data already empowers third parties to provide tools that extend Parliament's reach to a wide range of people around the UK. Parliament can use open data to project a degree of objectivity into an innovative and fast-changing online world. It already celebrates reusers of its data but there is an opportunity to do much more and to forge closer links with the companies and volunteers who add value and context and who are helping Parliament to meet its 2020 goals of being interactive, digital and understandable to everyone.

Background and approach

This report was delivered in seven weeks and evidence was gathered in three ways: desk-based reading, user interviews and analysis of tools built on Parliament data.

We read key works, such as the 2014 MySociety report,[1] the 2015 report of the Speaker's commission into digital democracy,[2] and user-research into public engagement with Parliament. Our interviews and discussions directed us to read more about parliamentary data structures,[3] and the need for public sector digital projects to pay due attention to back-end systems.[4]

We spoke with 25 people who worked with parliamentary data, including academics, journalists, charities, advocacy groups, parliamentary researchers, developers, open data enthusiasts, the data.parliament.uk team, and the National Audit Office. Of these discussions, 21 were with heavy users of parliamentary data. In all cases, we asked the following questions:

- *What data do you already use from Parliament? Can we make that easier for you?*
- *What would you like to do but can't? Is data.parliament.uk one of the obstacles to this?*
- *What do you already know about your users? Is there anything more you'd like to know?*

We gathered notes from these interviews but they are not published as part of this report. We received a number of additional email responses and recorded them alongside notes from our background reading.

Finally, we analysed a number of tools and services built on Parliament data. This included a comparison of TheyWorkForYou and data.parliament's user statistics, an investigation of tools created at accountability hacks,[5] and the use and documentation of existing websites including data.parliament.uk, hansard.parliament.uk, and parliamentlive.tv.

[1] Parliament's Online Services: A Strategic Review containing Strengths, Weaknesses and Recommendations, by MySociety, 2014.
[2] "Open Up!", Report of the Speaker's Commission on Digital Democracy, 2015.
[3] "United Kingdom parliamentary URL structure: change needed", Andy Mabbet, 2015.
[4] "Mind the GaaP: UK.gov needs to get a grip on digital", Kat Hall, The Register, 2016.
[5] See accountabilityhack.org.

Detailed recommendations

1) Focus on providing better data ahead of building more tools

The majority of interviewees use parliamentary data indirectly via tools such as TheyWorkForYou, PublicWhip, and WhoCanIVoteFor. Many developers use APIs from TheyWorkForYou in preference to those from data.parliament.

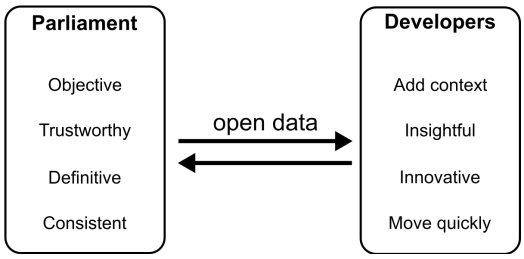
It is important to also see this as a success. Reuse of data in this way has enabled the creation of a wide variety of powerful and reliable tools that might otherwise not have existed. Importantly, almost every data user felt that Parliament had a unique and valuable role to play. As one participant said:

> *"Data.parliament.uk should focus on the things that no-one else can do. We want linked data in good formats with good APIs more than new tools"*

Parliament should continue to be the definitive source of data on what happens in Parliament. Users of this data can then add value by building easy-to-use tools quickly and by adding context, opinions, and insight to raw data.

Developers value the legitimacy that Parliament's data gives to their products. They are keen to link back to original sources wherever possible, but it is currently difficult or impossible for them to do so.

Open data is the link between Parliament, developers, journalists and the many people who use their tools and engage with their output. Adding Open Data Certificates to datasets is a small but important step to further increasing the trust that users have in the raw data.



There is a vibrant ecosystem around parliamentary data that is envied in other parts of the country and the world. To preserve this, **Parliament should focus its efforts on improving the data it provides** to developers in preference to building its own tools that compete with what already exists. The following recommendation sets out how we believe this should happen.

2) Deliver short-term improvements while building a long-term data model for Parliament

Updates to Hansard online, parliament.tv, and Hansard's search tool have been well-received but there remains an appetite for further improvement. Many data users feel that key datasets remain unavailable in a high enough quality or in the formats that they want.

We have heard multiple requests for improved access to the Register of Members' Financial Interests (RMFI), Select Committee membership data, lists of registered MPs' assistants and researchers, archived lists of ministerial statements, a live-feed of laid documents (documents formally deposited with the House of Commons),[6] and for better links within Hansard data.

Although these requests sound similar, it is important to distinguish between those that should be easy to fix and those that require longer-term investment.

Example 1: Members interests – an easy fix

There is a common perception that access to the RMFI continues to be unnecessarily difficult. It is one of the last datasets that TheyWorkForYou still scrapes from parliament.uk. Other users extract data from PDFs. The website membersinterests[7] provides an API to share their extracted version of the original data.

[6] See Guidance for laying papers before the Commons, Journal Office, 2016.
[7] See www.membersinterests.org.uk.

In all cases, the process of scraping adds costs to developers and reduces the end users' trust in the data. Scraped data is occasionally incorrect and it rarely links back to its definitive source at Parliament.

Developers want to access the RMFI as a well-categorised table for bulk download and as an API for querying. Failing this they would appreciate a public statement, further to the most recent in January 2016,[8] on what the obstacles are to achieving this.

Example 2: Bulk downloads limited at 500 rows – an easy fix

"I'd love a list of divisions as an up-to-date, bulk download but at the moment there's no easy way to get that."

Many users are unhappy that bulk downloads of datasets are limited to 500 row chunks, and find alternative methods of accessing many datasets limited in this way to be poor. This is not a new problem,[9] and seems to be largely the result of under-powered servers running the data.parliament.uk system.

These performance issues have an impact beyond arbitrary limits on bulk downloads. Several interesting internal prototypes that rely on data.parliament's API are affected. For example the MP Explorer[10] takes over a minute to return the 2723 votes that Andy Burnham has made since 2001.

This is poor performance for a modern API and stops these tools reaching wide use. It means that developers outside of Parliament replicate data locally or use second-hand providers of APIs such as TheyWorkForYou. This stops data.parliament.uk from understanding how many people use their data. It also reduces the likelihood that a link to the source Parliament data will be provided to the user.

Under-resourcing of critical parts of Parliament's data infrastructure was noted in MySociety's report[11] and this is clearly a challenging issue to address. Nevertheless, we think that **an investment in infrastructure to host data.parliament's APIs is essential and will deliver an improvement in both performance and user satisfaction**. Users must be able to download **full datasets** and have access to a performant API.

The need for a comprehensive data model, a multi-year project

[8] The Register of Members' Financial Interests, Jan 2016.
[9] See Just give us the data (Jul 2015) and Downloading data from data.parliament (Jan 2015).
[10] See MP Explorer by DDP.
[11] See page 10 of Parliament's Online Services: A Strategic Review containing Strengths, Weaknesses and Recommendations, by MySociety, 2014.

Developers and data users' frequently asked for stronger linking within the data that Parliament provides. Returning to the example of the RFMI this means that instead of free text in HTML files or PDFs the separate fields such as "Name of donor", "Address of donor", and "Donor status" are presented as separate columns within a table and as separate fields within the data returned by an API. Proper linking of data would allow users to ask questions like "how many MPs did <person or company> give money to?"

To provide richly-linked data we recommend creating a data model that spans Parliament. Creating such a comprehensive data model will not be easy. The difficulty is captured well in the following quote from a participant:

> *"The current data model is not formal or documented. There's a lot of duplication, with shadow identifiers in different parts of the data. The exception is members identifiers where there's a robust lookup table."*

This is an efficient way to resolve many of the hardest challenges set by users. **Work has already begun on formalising a data structure and providing more richly linked data.** [12] **We recommend that this work continues to be a top priority.**

The scale of this task should not be underestimated but nor should the value that this will bring to users, both external and to those within Parliament. Users requesting this improvement are well aware that this is a non-trivial task and they trust Parliament to get the data model right. They are mostly interested in the better data that this work will give them.

3) Parliament's tools and websites should run on its own open data

MySociety's 2014 report into Parliament's online services proposed that "separating the internet from ICT is no longer justified". We propose similarly that Parliament should adopt a more data-driven approach and end the separation of data.parliament.uk and parliament.uk. To show how this can work, we look to the way in which online petitions function today.

Example: petition.parliament.uk – breaking the barrier between data portal and website

1. At petition.parliament.uk/petitions the user is presented with a list of petitions.
2. Add .json to the end of the URL and the data powering that list is displayed in a machine-readable format.

[12] See UK Parliament Domain Model.

3. Select any of the unique identifiers from that list and add it to the end of the URL and you are taken to the page describing that petition. For example petition.parliament.uk/petitions/131215 displays details about petition 131215.
4. Add .json to the end of this URL and the data powering petition 131215 is displayed in a machine-readable format.
5. This data can then be used to create heatmaps of where the people who signed the petition live. Because these heatmaps are keyed by the unique ID of each petition it is easy to link to them.
6. Much more is possible. For example petition.parliament.uk/petitions?q=cats returns a list of petitions mentioning cats while petition.parliament.uk/petitions.json?q=cats returns the same list in a machine-readable format.

In this example we see that every web page is created from data that is directly available to the user. Each page is linked to a unique ID which allows third-party sites, or open data repositories such as WikiData, to refer and link to each petition unambiguously. Parliament should aim to emulate this data-driven approach to providing services.

The data model needed to describe Parliament is more complex than the data model needed to describe petitions. For the foreseeable future the data portal at data.parliament.uk will need to remain separate from the website. In this period, a number of users have suggested a compromise:

> *"If Parliament builds tools [such as the parliament.uk website] they should build them on open data sets and document how they work. That way they'll have to fix the open data first."*

We think is a good idea. A unique identifier system for Members of Parliament exists which would allow for parliament.uk URLs to be substantially improved. The larger changes to parliament.uk's URL scheme that Andy Mabbett describes in his blog[13] should be a long-term goal.

We recommend that an increasing amount of the existing content on parliament.uk is powered by data.parliament.uk APIs. This is already the case with many internal tools (such as MP Explorer) and this is appreciated by users. This approach should be extended to the parliament.uk website itself. If this is to be achieved the performance of the existing servers running data.parliament.uk will need to be improved significantly.

4) Work with the network to extend the reach of Parliament by doing what Parliament can't

We recommend that Parliament should:

[13] United Kingdom parliamentary URL structure: change needed, Andy Mabbett, 2015.

Support the network of data users and developers who create tools that extend its reach. Parliament should engage with developers to understand how its data can be improved. It should promote the best of these tools via the data.parliament.uk website, and offer to host guest blogs. It should consider creating a "Powered by data.parliament" digital badge so that users of parliamentary data can link back to the source. This would increase awareness of the data and APIs that are available.

Work with the network to understand and meet the needs of its end users. While Parliament should prioritise publishing useful and quality data over developing tools, understanding the user need remains important. Some user needs may not always be met by the market. Where this is the case Parliament may want to take action to address those needs, for example by running a hack or a challenge. Data users and developers like MySociety and Democracy Club can provide a useful proxy for understanding the end user need, in lieu of direct engagement.

Example: OKFN Germany's Prototype Fund,[14] the Urban Sustainable Development Lab,[15] and Manchester Digital Health Hack[16]

The Open Knowledge Foundation's prototype fund will assign €1.2m to open-source projects up to €30k in value.

The Urban Sustainable Development Lab operates internationally to pair technologists with public sector challenges and pay them for their time.

Manchester Digital's health hack was a two-day event supported by Public Health England's Health X innovation. It paired public health experts with paid developers for two days of prototyping.

Data.parliament, in collaboration with the National Audit Office have experience with hackathons through their accountability hack. A number of the people we interviewed attended these events and were positive about them.

Hack events are good for developing and testing new ideas but they are less good at shaping and involving users with long-term projects like the formalisation of a comprehensive data model for Parliament that we have recommended. As Parliament's data and APIs improve we think it is critical that developers can shape those changes and that they are helped to adapt to them.

> *"It's quite hard to find the time and money to update our tool to use the new APIs … I don't go to accountability hacks anymore because I rarely have time for that."*

[14] The Prototype Fund is a project of the Open Knowledge Foundation Germany, funded by the Federal Ministry of Education and Research: prototypefund.de/en.
[15] The Urban Sustainable Development Lab.
[16] Health Hack by Manchester Digital.2016 as part of the Health X initiative by Public Health England.

All previous reports have noted that Parliament has an unusually wide range of users. It cannot hope to consistently provide an excellent experience for all of these users. There is a temptation within Parliament to want to own the engagement it has with citizens. Whilst we recommend being aware of the wider user need, open data is a way of extending Parliament's reach via developers, activists, journalists, and politicians. We recommend that Parliament continues to shape its data strategy in this way.

Desired user journeys as tests of progress

In discussions with users of parliamentary data we asked whether there was anything that people wanted to do but currently couldn't or didn't know how to. Below is a selection of these requests in the hope that they will guide development and offer tests of progress. In many cases the questions may already be answerable, in some cases they may extremely difficult.

- "Can I ask the gender composition of a select committee?"
- "Can I create an automatically updating gantt chart of select committee membership just from open data?"
- "Can I create an automatically updating gantt chart of what an individual person has been doing just from open data?"
- "How would I see all papers laid with the House of Commons? Could I subscribe to a feed of this?"
- "In 2003 David Davies opposed some bill on the ground of civil liberties. We wanted to see who voted with him on this kind of issue since then so that we could contact them all to support an upcoming bill. We couldn't."
- **"**If I saw that an organisation had given evidence, how would I see all the evidence they'd ever given?"
- "If I saw that a company had given money, how would I see all the money they'd ever given?"
- "How would I see when a topic was debated and where? How would I skip to watch the video or jump to the full context in Hansard?"
- "I want to be able to download a huge list of votes as a CSV file."
- "I'd be really interested in the register of researchers, there's so much knowledge locked up in there."
- "I can subscribe to a list of written ministerial statements but there's no good archive."