CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

What do data portals do? Tracing the politics of online devices for making data public

Jonathan W.Y. Gray [ID]

Department of Digital Humanities, King's College London, London, United Kingdom
E-mail: jonathan.gray@kcl.ac.uk

Abstract

The past decade has seen the rise of "data portals" as online devices for making data public. They have been accorded a prominent status in political speeches, policy documents, and official communications as sites of innovation, transparency, accountability, and participation. Drawing on research on data portals around the world, data portal software, and associated infrastructures, this paper explores three approaches for studying the social life of data portals as technopolitical devices: (a) interface analysis, (b) software analysis, and (c) metadata analysis. These three approaches contribute to the study of the social lives of data portals as dynamic, heterogeneous, and contested sites of public sector datafication. They are intended to contribute to critically assessing how participation around public sector datafication is invited and organized with portals, as well as to rethinking and recomposing them.

Policy Significance Statement

Information policy involves not only laws and policy documents, but also digital infrastructures and online devices. Taking the case of open data portals, this article examines three approaches for studying the configuration of such devices: (a) interface analysis, (b) software analysis, and (c) metadata analysis. These approaches may support learning around how open data policy aims are translated into software and infrastructure, and may be used to assess effectiveness, surface limitations, and to inspire experimentation with alternative arrangements.

---

1. Introduction

The past decade has seen the rise of "data portals" as online devices for making data public. They have been accorded a prominent status in political speeches, policy documents, and official communications as sites of data-driven innovation, transparency, accountability, and participation. If data portals may be envisaged as sites where the relations between citizens, states, and data materialize (Ruppert, 2015), then how can we approach them as objects of study? How may they be empirically re-specified (Marres, 2012), and even put back together differently?

Drawing on research on a selection of data portals from around the world, open source data portal software, and associated online spaces and infrastructures, this paper explores three approaches for studying the social life of data portals as online devices: (a) interface analysis, (b) software analysis, and

CrossMark

(c) metadata analysis. These three approaches contribute to the study of the social lives of data portals as dynamic, heterogeneous, and contested sites of public sector datafication. As well as facilitating the empirical respecification of data portals as sites of material participation (Marres, 2012), critical engagements with data portal interfaces, metadata, and software may also inform the rethinking and recomposition of these infrastructures, as well as the relations and forms of life that they sustain. Device-sensitive approaches and methods often used to study private sector platform practices may also be drawn upon to evaluate the politics of public sector data innovation, and open up space for more inclusive, diverse, effective, and redistributive interventions.

2. Background

2.1. Data portal imaginaries, aspirations, and disappointments

What are data portals for? What is it hoped, promised, and expected that they will do? Different kinds of visions and promises associated with data portals can be traced on the web and online spaces. Some of the more bold and ambitious claims and aspirations may be found in archives of official announcements, political speeches, and press materials. Materials associated with the launches of the US's data.gov and the UK's data.gov.uk provide a sense of some of these earlier hopes, as well as subsequent disappointments. On the launch of the Data.gov, Vivek Kundra, the first Federal Chief Information Officer of the United States wrote: "The goal of Data.gov is to enable better decision-making, drive transparency, and help to power innovation for a stronger America."[1] About pages May 2009 state the goals of the site are to "encourag[e] innovative ideas," "make government more transparent," create an "unprecedented level of openness in Government," "strengthen our Nation's democracy" and to "promote efficiency and effectiveness in Government."[2]

On the launch of the UK's Data.gov.uk, Tim Berners-Lee, widely known as the inventor of the World Wide Web and founder of the World Web Foundation and the Open Data Institute, commented that it would: "unlock new ideas for delivering public services, help communities and society work better, and let talented entrepreneurs and engineers create new businesses and services."[3] The then government minister for Digital Britain contended that Data.gov.uk was "a cornerstone of how to strengthen the role of citizens and government, and recast the relationship between the two" and "a remarkable transformation in the availability of information to the public."[4] Then Prime Minister Gordon Brown said that the UK's data portal would contribute to "breaking down the walled garden of government, using technology and information to provide greater transparency on the workings of Whitehall and give everyone more say over the services they receive."[5]

These kinds of aspirations may be considered in relation to broader histories of open data as a malleable idea whose meaning is continually reconfigured in relation to shifting conceptions of governance and democracy (Gray, 2014). To the extent that these kinds of promises become socially shared and institutionally operative, they may be considered as "data imaginaries" (Rieder, 2018; Ruppert, 2018; Gray, 2018b; Lehtiniemi and Ruckenstein, 2019; Tupasela et al., 2020; Leonelli, 2021).[6] Different ideas about what data portals can achieve are encoded into more or less durable, shared understandings which may guide action. Data portals can be envisaged as part of what it means to deliver on broader promises or commitments to transparency, openness, participation, and innovation, such as when Obama's Open

---

[1] https://obamawhitehouse.archives.gov/blog/2009/06/08/data-transparency-datagov.
[2] https://web.archive.org/web/20090525091851/http://www.data.gov/about.
[3] https://webarchive.nationalarchives.gov.uk/ukgwa/20100304104714/http://www.cabinetoffice.gov.uk/newsroom/news_releases/2010/100121-data.aspx.
[4] https://www.theguardian.com/technology/datablog/2010/jan/21/government-free-data-website-launch and https://hansard.parliament.uk/Commons/2010-02-05/debates/10020534000001/LocalAuthorities(OverviewAndScrutiny)Bill.
[5] https://web.archive.org/web/20100324213649/http://www.number10.gov.uk/Page22897.
[6] Many of these draw on previous work on "social imaginaries" and "sociotechnical imaginaries" as shared visions of collective life and orientations towards the future (Taylor, 2002; Kelty, 2005; Jasanoff and Kim, 2015).

Government Directive stipulated that datasets should be registered on data.gov within 45 days; when the then G8 Open Data Charter included data portals as part of commitments to "open data by default"; or when members of the Open Government Partnership highlight open data portals as part of their national action plans.[7] In this sense data portals may also be understood in relation to performances of transparency (Currie, 2020).

Data portals also show up as part of commitments not just to openness but also taking action on issues. In the Obama archives Data.gov is mentioned in relation to better serving Asian Americans and Pacific Islander (AAPI) communities, protecting students from sexual assault, improving coastal planning, preparing for the effects of climate change and supporting civic participation through a National Day of Civic Hacking.[8] In the UK, one finds data.gov.uk turning up in relation to commitments to scrutinizing public finances with "armchair auditors" and cutting government spending, energy waste, more responsive public services, economic growth, improved transport, commercial innovation, tax transparency, public sector reform, and international open government commitments (Gray, 2014).[9]

As well as institutional data imaginaries, one can also look at how data portals are envisioned by activists, hackers, and civil society groups involved in making and using them. Data.gov and data.gov.uk are both run using the open-source CKAN software which preceded both portals.[10] The project started around 2007 as a "web application to manage listings of knowledge packages" hosted at the Open Knowledge Foundation, a non-profit organization dedicated to "promoting the openness of knowledge in all its forms."[11] CKAN stands for the "Comprehensive Knowledge Archive Network," an allusion to software repositories such as the "Comprehensive Perl Archive Network" (CPAN) or the "Comprehensive TeX Archive Network" (CTAN). These kinds of software projects serve as lists and registries of free/open-source software packages, libraries and components. They have been accorded a prominent role in free/open source developer and user communities, and reflect visions and practices of free/open source software development.

Interviews with contributors and materials associated with the development of the CKAN project suggest several ways in which the project has drawn inspiration from free/open source software development practices.[12] Firstly, the practice of software "packaging" or "componentization" is said to enable large, complex, and specialized software projects to be broken down into smaller components that can be used across multiple different projects and settings. Second, software registries can facilitate computer-assisted and semi-automated ways of working with data through accompanying command line tools—such as "apt get"—which support installation, package management, interfacing with other software packages and many other tasks. Thirdly, software registries facilitate distributed collaboration around the development of code—including basic social features as well as "versioning" to track and manage changes and contributions.

The transposition of these different valued features of free/open source software development led to the development of "data packages" and discussions of data "packaging" and "componentization" (complete with metaphors of shipping containers, logistics, physical production processes, and supply chains). A set

---

[7] See: https://obamawhitehouse.archives.gov/open/documents/open-government-directive, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex and https://www.opengovpartnership.org/?s=data+portal. Open Government Partnership members prominently including open data portals as part of their commitments and national action plans (2011–2022) have included Albania, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Colombia, Côte d'Ivoire, Croatia, Estonia, Georgia, Ghana, Greece, Guatemala, Honduras, Indonesia, Ireland, Italy, Jamaica, Kenya, Kyrgyz Republic, Latvia, Liberia, Lithuania, Moldova, Montenegro, Netherlands, North Macedonia, Papua New Guinea, Philippines, Portugal, Romania, Serbia, Sierra Leone, Slovak Republic, Republic of Korea, South Africa, Sweden, Tunisia, United States and Uruguay.

[8] https://search.archives.gov/search?query=%22data.gov%22&affiliate=obamawhitehouse.

[9] https://www.gov.uk/search/news-and-communications?keywords=%22data.gov.uk%22.

[10] https://ckan.org/.

[11] See: https://pypi.org/project/ckan/0.3/ and https://web.archive.org/web/20070316073641/http://www.okfn.org/.

[12] This includes interviews, exchanges and personal correspondence with CKAN contributors working on the core codebase and national data portals (~2016–2019), as well as reviews of documentation, repositories, web archives, mailing lists, blog and social media posts and other materials.

of libraries and command line features for data portals led the project to be initially described as "apt-get for data" at the leading Chaos Communication Congress in Berlin.[13] The project was framed as a contribution to enabling "collaborative data ecosystem" by facilitating contributions to a common set of datasets from a distributed set of actors, often with different projects and interests. For example, one "data wrangler" could download, clean up, reformat, or combine different datasets, and re-upload them, saving others from doing this same time-consuming or tedious data work.

Just as with software, the integrity of these kinds of incremental and collaborative knowledge development was envisaged to be underpinned by "versioning." A "versioned domain model" for the CKAN project—described as a "wiki-like interface"—was intended to enable contributors to keep track of changes not only in relation to datasets and their associated metadata, but also the data portal software itself. This was intended to enable more collaborative, transparent, and reliable "data work," as more people could work together around fewer shared components which would benefit from multiple improvements and increased scrutiny across projects (as per the moniker "many eyes make shallow bugs").

Archived web pages from the Wayback Machine can provide a flavor of the evolution of these earlier data imaginaries. The initial CKAN software was deployed on a website called "ckan.net" which in 2007 was described as a "register" (explicitly in the vein of the CPAN and "freshmeat" software projects) enabling anyone to find or create "open knowledge packages." By 2010, the front page says "this site makes it easy to find, share and reuse content and data, especially in ways that are machine automatable." By 2011 the site was renamed "The Data Hub" (thedatahub.org) and described as "the easy way to get, use and share data." By 2013 the site's front page (now at datahub.io) offered the following: "Give your data a home: Publish or register datasets, create and manage groups and communities."

The first public instance of the CKAN software may thus be considered as a kind of "infrastructuring" device—to assemble data publics and to facilitate distributed collaboration between them around an open knowledge "ecosystem" or "commons." This includes the aspiration that actors spread out across space, time, and in different settings could nevertheless contribute to enriching a shared pool of informational resources, facilitating a wide range of societal benefits. This genre of socio-technical imaginary can be read in relation to previous research on free/open source software development—including social studies of the values, imaginaries, and practices of free and open-source software communities (Kelty, 2008; Tkacz, 2012).

Perhaps not surprisingly given the scope, ambition, and implications of these various hopes and promises, data portals have not infrequently been met with disappointment, suspicion, and critique from many quarters. Some critique how data portals are implicated in broader problematic ideologies of transparency. For example, in a 2012 piece for *The New Inquiry*, Sarah Leonard writes of how government data portals and transparency activist groups such as Wikileaks both advocated the "virtuous data dump" which "swamps our capacity to make sense of it and becomes obfuscatory" and "makes it easy for states and social-media corporations to turn the tables and rationalize their relentless invasion of our privacy" (Leonard, 2012). Clare Birchall writes of how data portals and the open data initiatives with which they are associated call upon us to take on the "imaginary identity of the ideal data subject as a citizen–auditor–consumer–entrepreneur" in "info-capitalist-democracy," and suggests their spread may be viewed as a kind of "transparency imperialism" (Birchall, 2015).

Some raise concerns about how open data initiatives, and the data portals which play a prominent role in them, may have problematic effects—such as "empowering the empowered" if data portals and open data initiatives are not accompanied by further measures to enable "effective use" (Gurstein, 2011); serving as a Trojan horse for new public management (Longo, 2011); promoting the marketization of public services (Bates, 2012); and being plundered by "free market doppelgängers" which "follow civic hackers wherever they go" (Slee, 2012). Data users have also highlighted data portal shortcomings. In 2010 the *Guardian* reported developer community disappointment that postcode files were not made

---

[13] https://fahrplan.events.ccc.de/congress/2009/Fahrplan/events/3647.en.html.

available for non-profits following the launch of data.gov.uk, and in 2011 that data published on the site had not been useful or provided "value to people in their day-to-day lives."[14] A 2015 Nieman Lab piece on data.gov highlighted comments from journalists, transparency advocates, and developers that datasets were "old," "irrelevant" and infrequently updated, that important datasets were missing (e.g., election results), and the portal does not make data accessible for broader publics.[15] A recent review of literature on data portals suggests that literature is "almost universally critical, albeit often also constructive" and suggests an "expectations gap" between different aims and the "limitations, faults and unfulfilled promise of portals," including shortcomings related to machine-readable data, transparency impacts and broader accessibility (Davies, 2021).

2.2. Data portals as online devices

In light of these various expectations and disappointments, how can data portals be studied? Against which and whose expectations should data portals be evaluated? How can we study not only what has been said about them (whether statements of hope or frustration), but also their material organization and computational affordances? As recent reviews attest, there are many ways of studying open data portals from a wide variety of disciplinary perspectives—including computer science, information management, political science, public administration (Davies, 2021). A not insignificant body of literature has been dedicated to developing assessment frameworks (Alexopoulos et al., 2017; Machado et al., 2018) and evaluating data portals *en masse* with various methods and measures from metadata quality (Neumaier et al., 2016) to usability scoring (Nikiforova and McBride, 2021).

This paper focuses on what social and cultural research grounded in fields such as science and technology studies and internet studies might contribute. Methodologically other studies in these fields have undertaken interviews (Boucquey et al., 2019) and ethnographies (Ratner and Ruppert, 2019) —as well as exploring how ethnographic approaches can be combined with digital methods (Burns and Wark, 2020). In complement to such approaches, this article examines ways of studying data portals as online devices, inspired by the approach of digital methods studies in this area, such as on the role of web trackers and coding platforms in news work (Bounegru, 2019) and the temporalities of "real-time" web devices such as search engines and social media platforms (Weltevrede et al., 2014). These studies draw on and are grounded in research attending to devices in science and technology studies (STS), in particular work on how devices acquire political capacities and take part in politics and public life (Marres, 2011, 2012; Marres and Lezaun, 2011); how digital devices are involved in producing and performing contemporary sociality through their reconfiguration of knowledge spaces (Law and Ruppert, 2013; Ruppert et al., 2013); and "market devices," "material and discursive assemblages that intervene in the construction of markets," such as pricing models, trade protocols and economic indicators (Callon et al., 2007).

Research on "transparency devices"—such as transparency agendas, conflict maps, and administrative protocols—proposes that "sociotechnical devices should be understood not simply as 'gadgets' that do things but as complex and unstable assemblages that draw together a diversity of people, things and concepts in the pursuit of particular purposes, aims, and objectives" (Harvey et al., 2013). Ruppert proposes that devices associated with the UK's transparency agenda can be understood as part of a process of "enacting the transparent state" (Ruppert, 2015). Also highly relevant to the approach of this article is research on "transparency as political device"—looking not just at "talk about transparency," but also "localized and technical" practices such as displays of accounts, the conventionalization of conduct through guidebooks and the gathering of witnesses (Barry, 2013). Similarly, Hansen and Flyverbom propose to "look carefully at the contingencies of transparency, and at how it is produced" including through what they call "disclosure devices": mediating technologies involved in the production of

---

transparency, such as due diligence, rankings, and big data analysis (Hansen and Flyverbom, 2015; Flyverbom, 2016).

Device-sensitive approaches have also been taken up in the study of policy—including with the notion of "policy devices" (Hirschman and Berman, 2014; Frankel et al., 2016; Gray, 2017). For example, Hirschmann and Berman look at the complementary notion of "policy devices" which they suggest includes: "the wide variety of sociotechnical tools that help policymakers see and make decisions about the world," giving examples of GDP, unemployment rates, and censuses, arguing that "policy devices shape what we attend to" (Hirschman and Berman, 2014).

To consider how data portals can be studied as online devices, therefore, implies looking at them as dynamic, contingent, shifting, and unstable socio-technical assemblages which are conferred with political capacities, which take part in political life, which script collective action and relations and which shape how things are made intelligible. STS work on devices emphasizes how they may be understood as "patterned arrangements" which are have "social lives" in the sense that they are both shaped by the social and format social relations (Law and Ruppert, 2013). In what follows, I explore how data portals may be studied as online devices with such sensibilities. While many previous studies of data portals have involved large-scale data analysis, scraping, and analysis of interface features, the approaches below aim to repurpose online materials and attempt to re-configure online devices for the purposes of interpretive research on digitally mediated sociality and culture (Rogers, 2013; Marres, 2017). They are proposed as methodological entry-points for the study of data portals, illustrated through engagements with a selection of portals from around the world as well as the open source CKAN software which many portals use, notably data.gov and data.gov.uk which are two of the earliest and most widely mentioned examples in research, practitioner and policy literature.

3. Studying Data Portals as Online Devices

3.1. Interface analysis

Data portals may be studied through the interfaces that are presented upon visiting them. Such interfaces may be considered to configure users (Woolgar, 1990) and guide conduct (Light et al., 2018). Software studies and new media scholars have examined how interfaces script and order interactions and relations between hardware, software, data, and users (Pold, 2005; Cramer and Fuller, 2008; Andersen and Pold, 2011, 2018; Galloway, 2012; Chun, 2013; Hookway, 2014). Interfaces can be understood to be political in the sense that they produce particular kinds of users (Chun, 2013) and can be viewed as "zones of contestation" where human-machine relations are negotiated (Hookway, 2014). Interfaces may be envisaged to infrastructure publics (Dantec and DiSalvo, 2013) and format broader societal engagement with datasets (de Mourat et al., 2020). Dourish contends that interfaces can be viewed as "objects that structure and organize the world around us in meaningful ways" (Dourish, 2007).

How to study data portal interfaces? Drawing on a common distinction in web development, Richard Rogers distinguishes between the "front end" and "back end" of information politics—with the latter including "information formatting" and what is presented to the user and the latter including "information retrieval" techniques such as "search engine logics" (Rogers, 2004). We may start by looking at the interfaces which are presented to data portal visitors—the presence, absence, and ordering of different kinds of interface elements. This can be explored through repurposing the method of "reverse wireframing."[16] Wireframing is a practice of establishing and arranging the basic elements of a page for the purposes of user interface design, development, and testing. It is often used to "mock up" the composition of various elements—whether with wireframing software or through paper prototyping. "Reverse wireframing" can be used as a technique to analyze and decompose the basic elements of a web

---

[16] This approach is inspired by "The People's Dashboard" project at the Digital Methods Winter School 2015, which examined interface elements of social media platforms: https://wiki.digitalmethods.net/Dmi/PeoplesDashboard.

Figure 1. *Front page of data.gov—original (left), obfuscated text (middle), wire frame outline (right).*

page. For example, Figure 1 shows a reverse wireframe of data.gov, created using the "wirify" bookmarklet which lets users "turn any web page into a wireframe in one click."[17]

On the front page of data.gov one can see (from the top): a top menu bar (including a link to the "catalog" powered by CKAN, data topics, developer pages, and contact form, a link to a feedback form, a one-line summary ("the home of the U.S. Government's open data"), a count of total datasets ("search over 335,221 datasets"), a search bar with rotating search suggestions ("Monthly House Price Indexes," "Health Care Provider Charge Data," "Credit Card Complaints," "Manufacturing & Trade Inventories & Sales," "Federal Student Loan Program Data"), highlighted datasets, embedded recent blog updates, embedded recent Twitter posts and a footer with links.

Drawing on digital methods research on doing web histories with the Wayback Machine (Rogers, 2017), monthly snapshots of the front page of Data.gov.uk tell a story about how the site has developed (Figure 2).

Data.gov.uk is comprised of a shifting set of interface elements. It is described as being in "perpetual beta" (a term associated with Web 2.0 as formulated by Tim O'Reilly) in which "best practice, technical development, innovative use of data, and citizen-centric politics combine to drive data-release programs" (Shadbolt et al., 2012).[18] In this sense, the changes in Data.gov.uk are envisaged as part of a process of experimentation to generate societal value through the re-use of government data. From its launch in 2009 to 2022 there have been at least nine major interface changes, from private beta to trying many social web

---

[17] https://www.wirify.com/.
[18] https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=5.
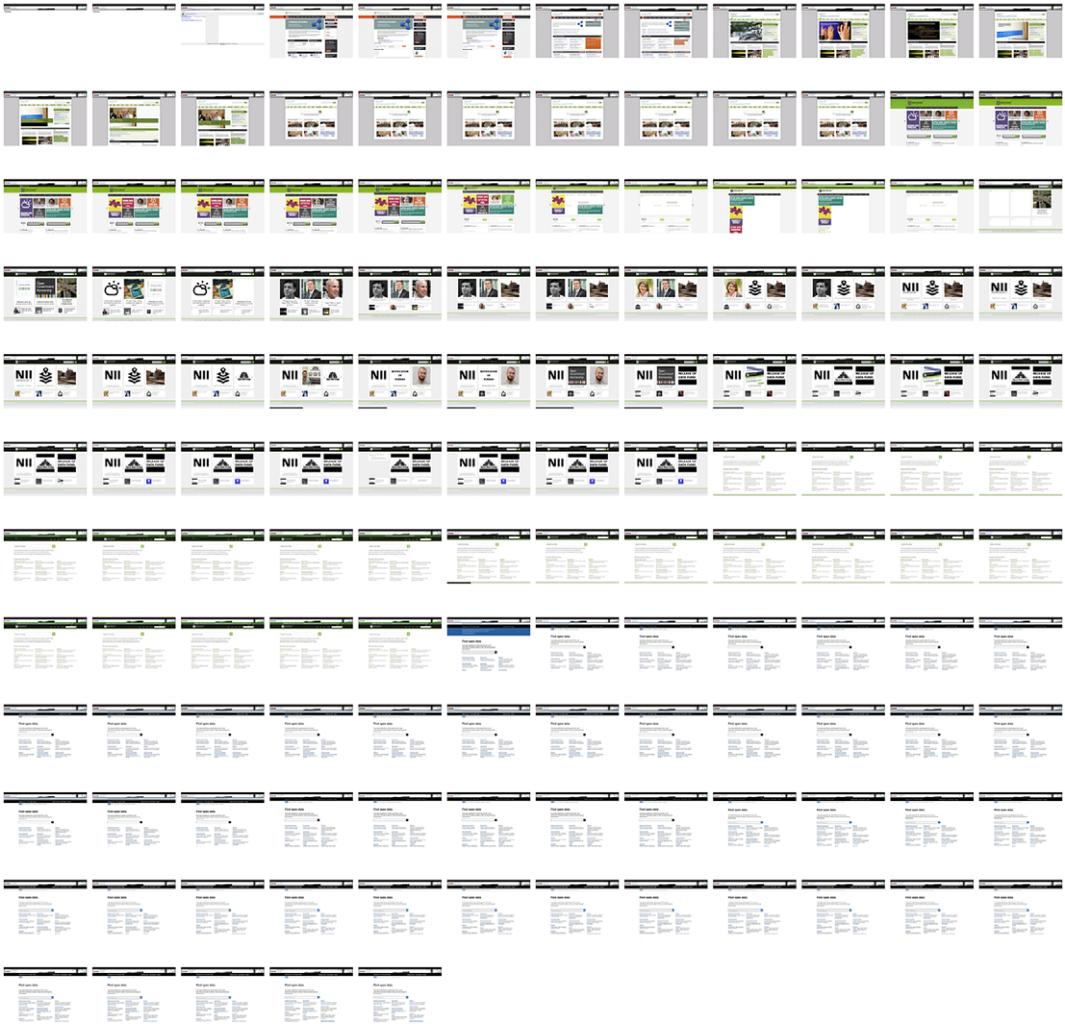
Figure 2. *Snapshots of data.gov.uk front page by month according to captures from Internet Archive's Wayback Machine (2009–2022). An animated version can be found at: https://vimeo.com/720658997.*

features to paring back to a simple search and browse interface with the official Government Digital Service theme:

- Jul 2009-Dec 2009: "Private beta." Empty page with "Welcome" followed by redirection to private "UK Government Data Developers" Google Group.
- Jan 2010-Sep 2010: "Unlocking innovation." Semantic web innovation and web 2.0, with SPQARL endpoint and site sections for "Apps," "Ideas," "Forum," "Wiki" and prominent social media links.
- Oct 2010-Nov 2010: "Transparency is at the heart of this Government." Rebrand with same social web elements. Linked data moved down, apps moved up. Highlighting of spending data.
- Dec 2010-Jun 2011: "Opening up government." Transparency has been taken off the main menu. Prominent texts on "Who is who in government and where does the money go?" and "Facts, figures, apps and more." Showcasing apps. Guidance for data publishers. Call for developers to request datasets.
- Jul 2011-Jun 2012: "Enter your postcode." Location-based search box to find "tools and information about your local area using government data."

- Jul 2012-Oct 2013: "Latest datasets, blogs, forums, and ideas." Rotating carousel, search boxes for data and content, total dataset count (from 8600 to 10,200).
- Nov 2013-Jul 2016: "National information infrastructure." Revamp coinciding with UK hosting Open Government Partnership. National information infrastructure and Open Government Partnership UK National Action Plan. Data portal as transparency showcase. Top menu simplified to "data," "apps" and "interact." Menu items for forum, data requests, and linked data are removed.
- Aug 2016-Apr 2018: "Search for data." Main items on page are big search box, plus "browse data by theme" with twelve suggested topics—the order of which suggests how role of state is envisioned (business and economy, crime and justice, defense, education, environment, government, spending, health, mapping, society, towns, and cities and transport).
- May 2018–May 2022: "Find open data." Rethemed to match Government Digital Service theme. Emphasis has shifted from process ("search for data") to outcome ("find open data"). The menu items for "data," "apps" and "interact" have been removed and replace with "publish your data," "documentation" and "support" aimed at government departments wishing to publish data, indicating a shift towards supporting data publishing.

While some researchers suggest that portals "mature" towards incorporating more web 2.0 features (Alexopoulos et al., 2017), web snapshots of data.gov.uk indicate a process of digital de-socialization and de-experimentalization: the removal of social features and a shift from broader engagement to more minimal site focused on search, browsing, and publishing. From its founding aspirations to elicit ideas, interests, communities, and contexts of potential data usage, the portal transitions towards a pared-down official directory.

As well as looking at interface elements and histories for individual portals, patterns may be detected by looking at the interfaces of multiple portals. Taking screenshots of a selection of 87 national data portals from a register of portals around the world triangulated with other sources and queries, one can observe several shared interface design patterns (Figure 3).[19]

Taking this as a way to read portal aesthetics and interface conventions through "non-reductive visualization" (Manovich, 2008) one can distinguish the prominence of white search bars; icons indicating different data topics; graphs, charts, and visualizations; and content boxes with photos, blog posts, and other news and updates. With a few more colorful exceptions, the dominant palette appears to be composed of shades of grey and blue, which may be taken to suggest a muted and cooler informational aesthetic (Liu, 2004), an atmosphere of neutral, official display.[20]

Manually coding the different types of interface features of these 87 portals indicates which features are more and less common, and which are shared by which portals. In doing so we can look beyond individual portals toward shared conventions and styles of configuration (Figure 4).

As can be observed in the network graph, amongst the most common and widely shared interface elements are search interfaces, taking search boxes as a way to mediate between institutional informational resources and visitor interests: "What are you looking for?" (Figure 5). While in some cases the search boxes are tucked into navigation menus, in others they have been conspicuously placed them in the center of the page—which may be taken as a kind of interface "Googlization," following the search engine's pared down landing page.[21]

Many of these portals also feature spreads and grids of data topics, such as "health," "environment," "science and technology," "energy," "economy," "justice," "transport," "population," "agriculture,"

---

[19] National data portals were taken from https://dataportals.org/ combined with other triangulated lists and queries for variations of data portal, data catalog and open data combined with country names in order to double check for portals from countries not listed on dataportals.org. See: https://doi.org/10.6084/m9.figshare.21740051.v1.

[20] This colour composition is confirmed by Tineye's MulticolorEngine colour extraction tool, which suggests the front pages are ~90% shades of white and grey, and ~ 7% shades of blue with other colours less than 1% each: https://labs.tineye.com/color/.

[21] This was proposed as part of a group project on "What Do Data Portals Do?" at the Digital Methods Winter School 2017 at the University of Amsterdam, co-convened with Sabine Niederer: https://wiki.digitalmethods.net/Dmi/WinterSchool2017.
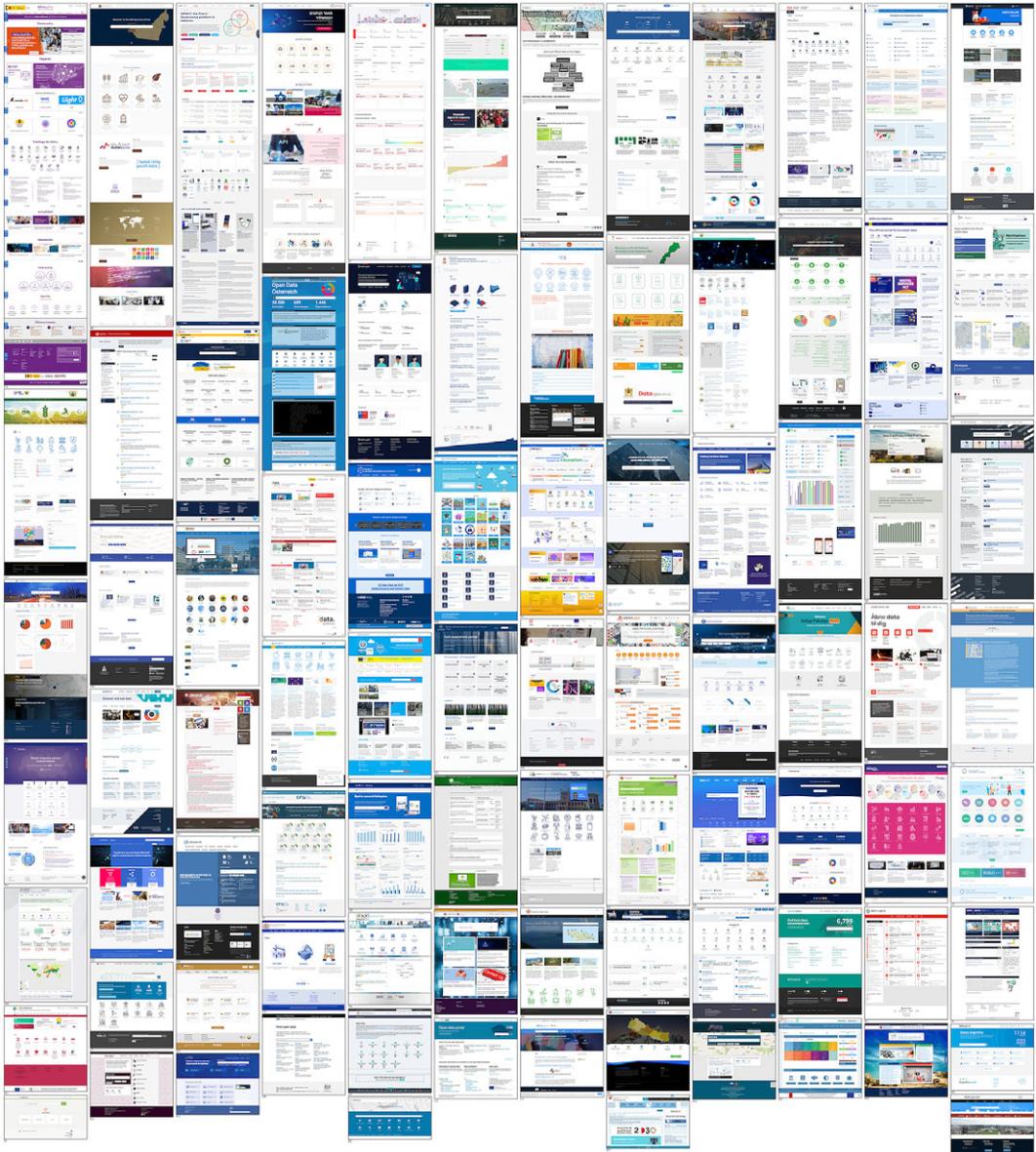
Figure 3. *Screenshots of front pages of a selection of 87 national data portals.*

"education" and "culture" (Figure 6).[22] The arrays of visual icons are reminiscent of the work of Otto and Marie Neurath's Isotype Institute: abstracted, encyclopedic, simplified, minimal, accessible visual depictions of different areas of institutional work (Neurath and Kinross, 2009). These topical directories complement search interfaces as an entry-point to datasets.

As well as being read against longer histories and aesthetics of the informational and informational graphics the interfaces evoke core elements of graphical user environments, operating systems and apps—with iconography serving as topical interfaces to data collections. The topicalization of data also

---

[22] These were found to be some of the most prominent topics according to research undertaken with student researchers as part of the King's Undergraduate Research Fellowship (KURF) scheme in Summer 2021.

Figure 4. *Bi-partite network showing interface features (green) of 87 national data portals (red). Spatialized in Gephi with ForceAtlas2 algorithm (Jacomy et al., 2014).*

discloses the curatorial role of portal-makers in displaying official "aspirational ontologies" (Marres, 2012) of collective life and collective concern. At first, these may be read as areas where data is most widely available, as an indicator of datafication—but it appears that for many portals these front page topics do not necessarily correspond directly with the topics or categories of data most prevalent amongst published datasets.[23] This is most evident in the case of empty data topics: topics with no datasets associated with them—such as the Mongolian open data portal's thematic placeholders for "tourism," "environment" and "foreign relations," each with "0 data components" at the time of writing. Perhaps they should rather be taken as "official accounts of reality" (Rogers, 2004), an indication of topics deemed to be of broader societal interest by governments and data portal designers, presented as an invitation to explore key areas of collective life as accounted for through the state and public institutions. Data portals serve as graphical user interfaces (GUIs) between data publishers and data publics and data users, with curated data topics serving as designated entry-points to informational resources of the state.

---

[23] This was a finding of the "Is this Open Data Portal accessible?" group project at the Digital Methods Summer School 2017: https://wiki.digitalmethods.net/Dmi/OpenDataPortalAccessiblility.

Figure 5. Selection of search prompts from front pages of national data portals.

Another prominent feature of data portal interfaces is usage stats related to data portal activity such as total datasets, APIs, downloads, site visitors, downloads, discussion thread numbers, counts of contributing bodies, and numbers of blog posts, apps, visualizations and data requests (Figure 7). While there are also statistics pertaining to the life of the countries and regions that the portals contain data about—sometimes reminiscent of "at a glance" style summaries, dashboards, and country profiles that may be found on web pages of international organizations and the CIA World Fact Book—portal stats focus on the quantification and display of how data is made public. As historical and social studies of quantification explore the making of official numbers (Desrosières, 2002; Espeland and Stevens, 2008; Bruno et al., 2016), portal stats may be understood as the making public or socialization of official numbers, displaying what is entailed in publishing data, indicators of success, traces of the work of cajoling government departments, cleaning and uploading datasets, publicizing portals, engaging with users, making applications, organizing events, and so on. Evident in these displays is the sense in which portal users may be those interested in statistical summaries of the work of opening things up—which may also include government departments, other countries, international organizations, funders, NGOs, civil society groups, and other professional users.

There are also various kinds of showcases with featured apps, app catalogs, app galleries, re-use stories, case studies, re-use studies, impact stories, and other kinds of content related to how data has been re-used. These sometimes mimic the interfaces of app stores such as Apple's app store and Google's play store, such Data.gov.uk's "latest apps" section, positioning data portals as a kind of digital infrastructure upon which innovators and entrepreneurs make things with data, perhaps in the spirit of Tim O'Reilly's
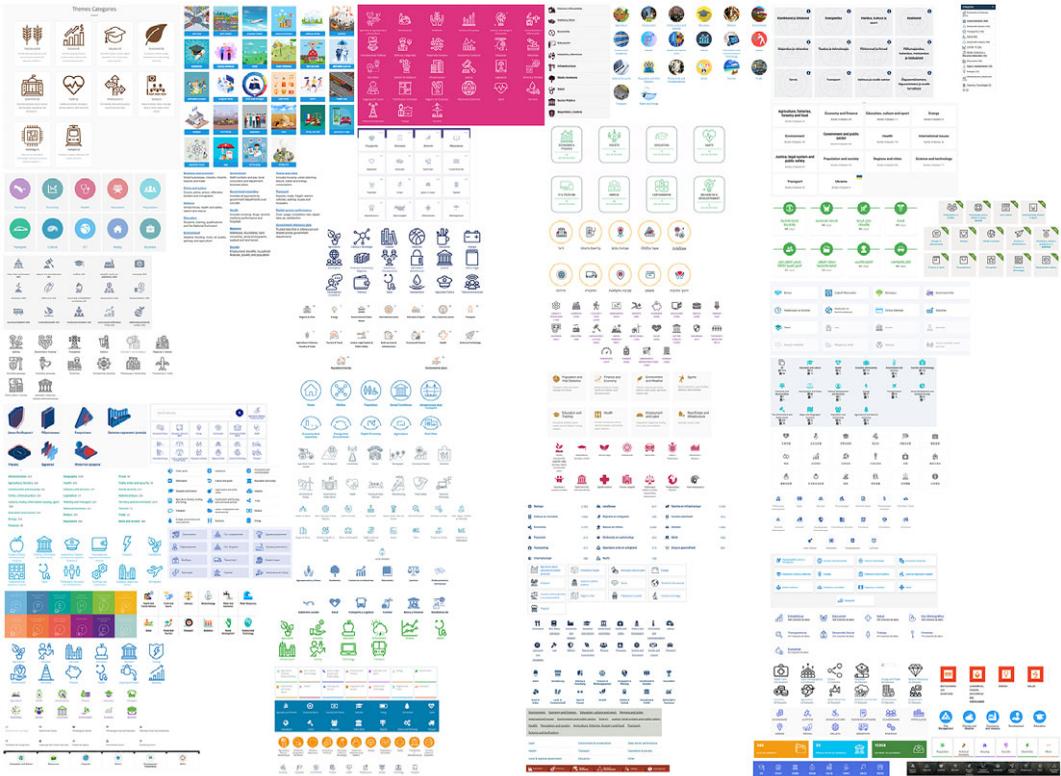
Figure 6. *Selection of data topics from national data portal interfaces.*

"government as a platform" (O'Reilly, 2010; Morozov, 2013). Data.gouv.fr contains a "Réutilizations" (re-uses) section, with over 3000 examples of how French open data has been used, browsable by topic, with each re-use case listing the datasets which were used, metadata, brief descriptions, and images and buttons inviting users to add further re-use cases, contact re-users, contribute to public discussion threads about the case, or embed or link to the project. The EU's data portal data.europa.eu contains use cases that can be filtered by region and sector, linking to brief case studies for each project, as well as over 400 "studies" on data re-use best practices, impacts, value estimates, event reports, roadmaps and "community reports."[24] Some portals such as Bangladesh's indicate activity by showing popular searches and datasets.[25]

Invitations to make with data also explicitly address developers, including through dedicated pages and spaces for developers (e.g., in portals from Bahrain, Japan, and Singapore). Many portals include APIs or application programming interfaces for both the data portals (e.g., to program requests to portals) as well as for particular datasets (e.g., "real-time APIs" for bus departures and arrivals, rainfall, river levels, air pollution). Some portals offer SPARQL endpoints (e.g., data.europa.eu) to facilitate machines to retrieve linked data. Hence invitations to make with data are not just to take data files as bundles of facts, but also to connect machines with information infrastructures which are part of the official organization of things. Data portals are not only "human-machine interfaces" (Suchman, 2007) but also machine–machine interfaces, inviting involvement from "non-human Users" (Bratton, 2016).

---

[24] https://data.europa.eu/en/impact-studies/studies.
[25] http://data.gov.bd/.

Figure 7. *Selection of activity stats from national data portals.*

3.2. Software analysis

The study of data portal software offers a complementary perspective on processes of making data public. Software codebases may serve to indicate the conceptual, logistical, and organizational work that goes into infrastructuring public data, including through software extensions that serve to adjust and extend code to suit the settings of deployment, such as to support metadata alignment, translation, customization, interoperability and support for specific data types or file formats. Data portals also embody different kinds of logistical aspirations, from making data "programmable" (Helmond, 2015), to command-line based packaging, to producing read-write "ecologies" and "ecosystems" of public data. Comparisons of different combinations of extensions may suggest different styles of configuration. The exploration of "platformized coding" on GitHub (Bounegru, 2019) provides a behind-the-scenes look at the various forms of collaboration and transnational coordination involved in adapting data portal software, as well as the organization of infrastructural work, maintenance, and care involved in making data public.

The CKAN codebase on Github gives a sense of how the project has developed from 2007 to the present, with over 23,000 commits from over 390 named users (Figure 8).[26]

In 2007 the project is described as a "web application to manage listings of knowledge packages," modeled on software registries.[27] By 2011 it is described as "opensource data hub and data portal software."[28] What is initially set up as a project to support collaborative knowledge development and

---

[26] https://github.com/ckan/ckan.

[27] https://github.com/ckan/ckan/tree/5bd35ffe3e23b3401581d59e4df45cfcc8adf223.

[28] https://github.com/ckan/ckan/tree/f6d6dd28e61f8778e7d813996542b13c0967d1e3.

Jan 14, 2007 – Jul 4, 2022

Contributions: **Commits** ▾

Contributions to master, excluding merge commits and bot accounts
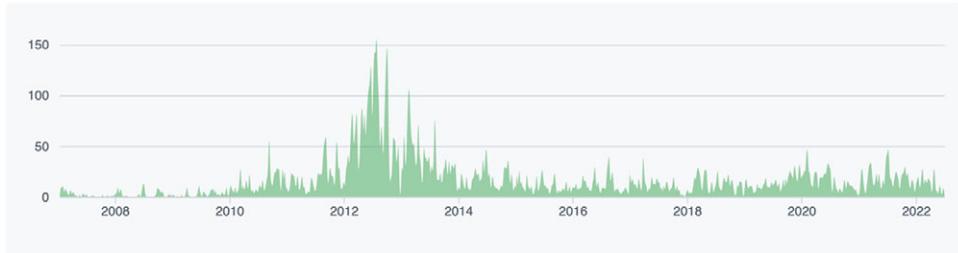
Figure 8. *GitHub contributor graph for ckan/ckan repository (2007–2022).*

described as a cross between a wiki, listing, and software package index is translated and adapted to support the publication of open government data. While earlier versions present a more open-ended and social webby domain model of "users," "packages," "tags" and "groups," later versions suggest a more data-focused and professionalized model of "datasets," "resources," "organizations," "authorization," and controlled "tag vocabularies." CKAN developers said that the project had to negotiate between this earlier vision of collaborative knowledge development and the needs and requirements of governments using the project.[29] Some government users did not have close contact with the main CKAN project development team and used the code independently. On the project mailing list one US government developer asked: "Is there a reason that forking is so pervasive in the CKAN community?." The project has over 1800 forks on GitHub—not counting many other forks and working copies of the code base hosted outside of the platform. One way the project team sought to coordinate different the needs of different users and keeping a shared code base was to enable a CKAN "core" code base and an "ecosystem" of themes and extensions.

The translation and adaptation of CKAN code can be detected in data.gov.uk's GitHub account, indicating different kinds of work involved in opening up official data—such as customizing a copy of the CKAN code base, harmonizing and exchanging metadata, enabling data schemas, optimizing search services, letting users browse and preview data, scheduling, and testing (Figure 9). Repository activity tells us about the temporality of portal software development. A flurry of new repositories are present in 2016–2017 as the portal is simplified (corresponding with the interface analysis above), after which new repository creation slows down as development settles and stabilizes—and the "perpetual beta" appears somewhat less lively.

Not all data portals have public GitHub repositories, and not all portal code repositories are actively used. For data portals using the CKAN software and with the API enabled, one can obtain details of the software version and extensions enabled. Taking a selection of data portals around the world using the CKAN software and their extensions as listed through the API provides a complementary perspective on data portal configuration (Figure 10).

In the graph, one can see which extensions are most commonly shared in the center (e.g., "stats" and "datastore") as well as portals that have more unique extensions and customizations which are not shared by other portals towards the edges (notably data.gov and data.gov.uk towards the bottom of the graph). Some clusters of extensions are associated with similar kinds of functionalities, such as metadata harvesting and exchange (lower center, around "harvest") and data and document previewing (top left, around "text preview"). These may be taken as an indication of convergence and variations in data portal software configuration.

---

[29] Interviews with CKAN developers and project contributors.
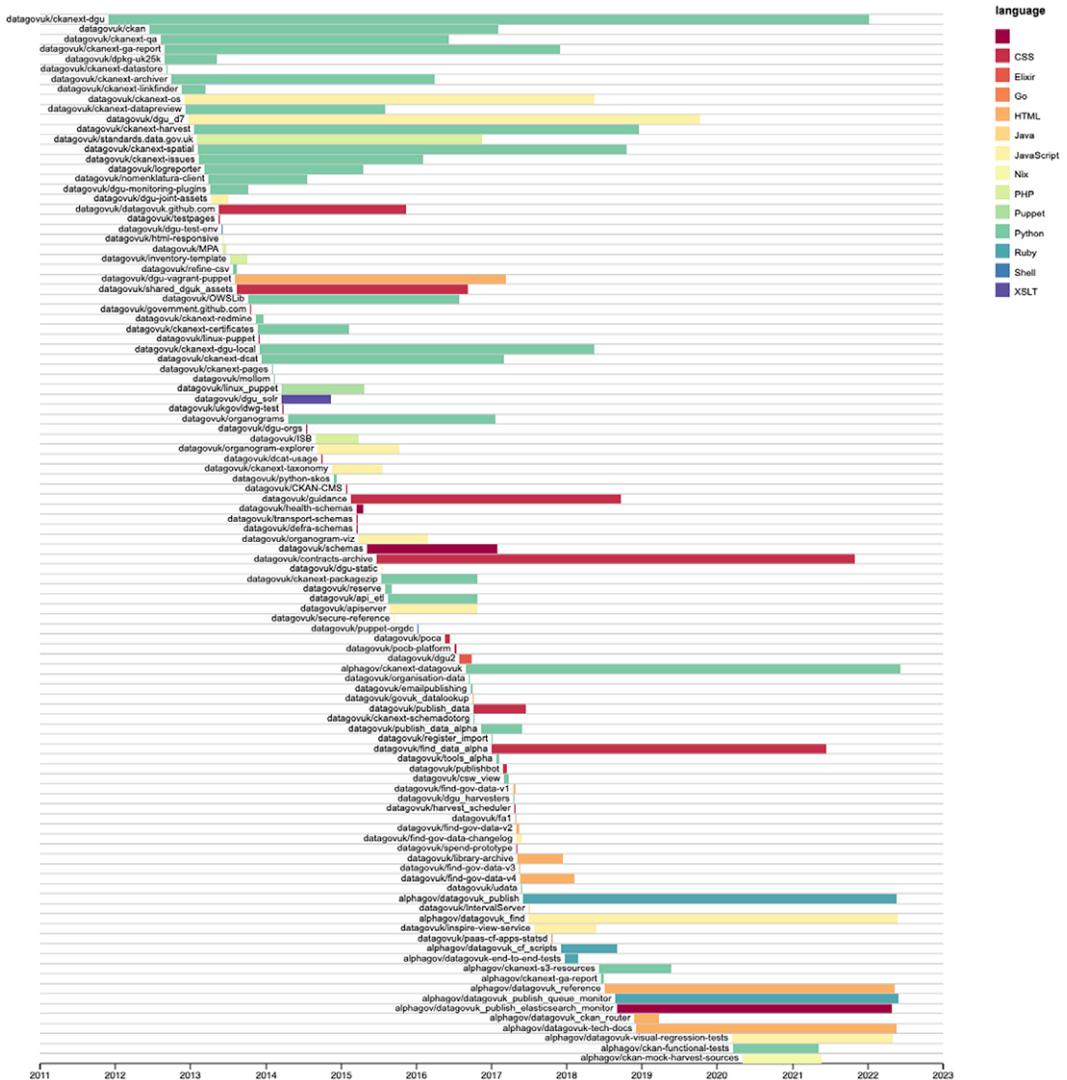
Figure 9. *Github repositories associated with Data.gov.uk (2011–2022).*

Data portal software indicates the scripting of various relations, entities, actions, operations, and activities involved in making data public. In the case of data portals using CKAN software one can observe the transition from a service modeled on a free/open-source software package manager to an extensible and customizable Wordpress-like "data management system" for government data publication with support for different aspects of behind-the-scenes data work involved in making informational materials into a re-usable resource.

3.3. Metadata analysis

Data portal metadata has been widely used as the basis for reports on the effectiveness of data portals and open data initiatives. Some researchers assess metadata quality (Neumaier et al., 2016; Nogueras-Iso et al., 2021). Data.europa.eu provides metadata quality pages including ratings of "findability," "accessibility," "interoperability," "reusability" and "contextuality," building on elements of the Data

Figure 10. *Bi-partite network graph showing data portals (red) and the extensions that they share (blue). Created at Digital Methods Winter School 2017 using the Gephi software and spatialized using the Force Atlas 2 graph layout algorithm.*[30]

Catalog Vocabulary (DCAT) standard.[31] As seen in the previous section, data portal software often supports data publishers to create, modify and work with metadata about government datasets. Data portals often don't store datasets directly, but link to external pages or sites where data may be downloaded—which has the implication that data portals may focus more on metadata curation and linking than on dataset publication per se. If data portals are aggregation mechanisms, then metadata can tell us both about the resources which are aggregated, as well as processes and activities involved in gathering and making data public.

What kinds of metadata do data portals provide and how can this be used to study them as online devices? The DCAT vocabulary aims to "facilitate interoperability between data catalogs published on the Web" through providing a model for data portal metadata—including for catalogs, datasets, data services, resources, publishers, and relations between them (Figure 11).

The Data Portal Explorer tool was created in order to gather and work with metadata from data portals, with an initial focus on those running on CKAN.[33] As per its original vision based on software package management, CKAN was designed not only for human data users, but also to enable machines to gather

---

[30] A finding from the "What Do Data Portals Do?" project at the Digital Methods Winter School 2017, University of Amsterdam, co-convened with Sabine Niederer: https://wiki.digitalmethods.net/Dmi/WinterSchool2017.

[31] https://data.europa.eu/mqa/methodology?locale=en and https://www.w3.org/TR/vocab-dcat-2/.

[33] Created with Jose Miguel Monteiro Vieira at the King's Digital Lab. More details at: https://publicdatalab.org/projects/data-portal-explorer/ and https://dataportalexplorer.readthedocs.io/en/latest/.

Figure 11. *Diagram showing DCAT 2 metadata model, W3C.*[32]

and work with datasets and associated metadata. Data portals thus facilitate the programmability of public data, whether for individual data users or for interoperability with online services, apps, and platforms.

Informed by work on the histories and sociologies of quantification (Desrosières, 2002; Espeland and Stevens, 2008; Bruno et al., 2016) analyzing data about datasets may provide both historical and comparative insights into the politics of public sector datafication: which kinds of topics and concerns appear when, how they are "parameterized" (Gray, 2018b), and which receive attention from data portal makers and users. For example, looking at metadata on topics and organizations on data.gov.uk (Figure 12), one can see many datasets related to environment and mapping from a handful of organizations who appear to contribute the greatest numbers of datasets (e.g., United Kingdom Hydrographic Office, Natural England, Environment Agency) —while also seeing that government spending data has the greatest number and diversity of different organizational contributors by topics, reflecting the liberal-conservative

PRIMARY THEME

UNIQUE CONTRIBUTORS

NAME OF THE ORGANISATION
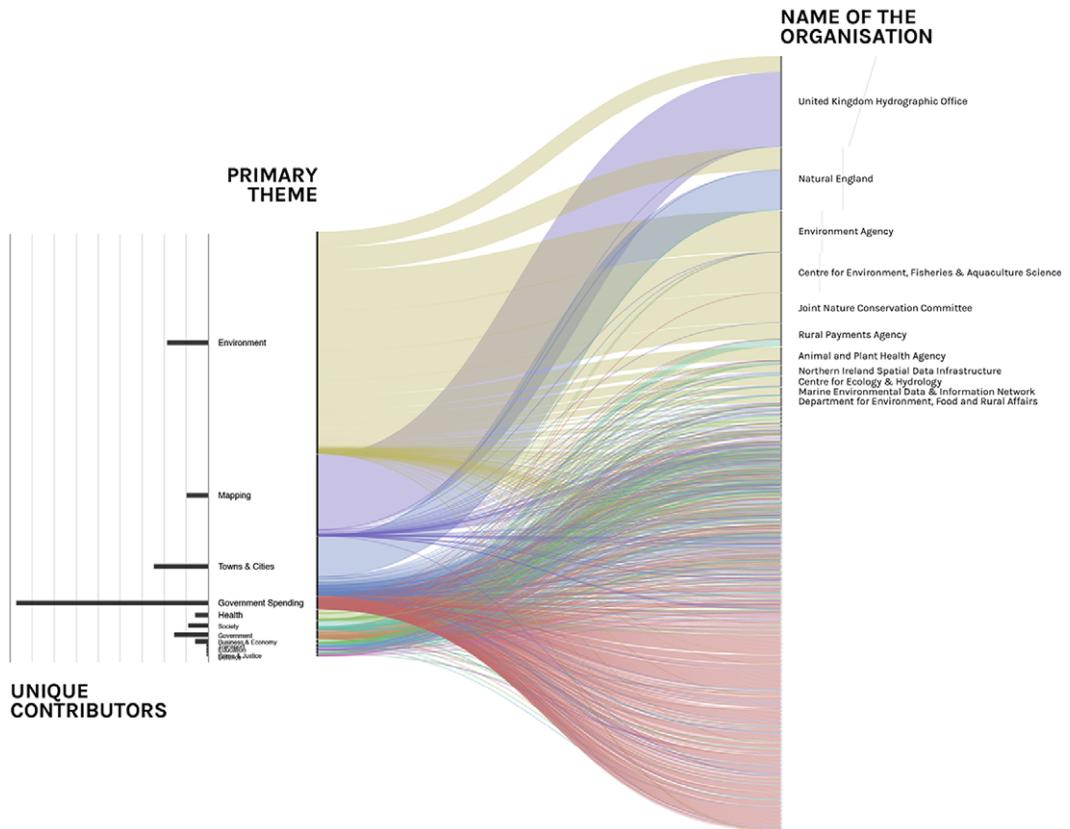
Figure 12. Alluvial chart showing relation between number of datasets added to Data.gov.uk by data "themes" or topics (left) and public sector entities (right), created at Digital Methods Winter School in Amsterdam, January 2017.

coalition government's spending transparency commitments (Gray, 2014). The data that ends up on data portals is a mix of what is most widely available, and what has been politically prioritized.

As well as broader data topics, datasets can also be tagged with more granular and specific tags. One can observe tags related to particular data collection methods ("census-2016" on data.gov.ie, "in-situ-laboratory-instruments-profilers-sounders-acoustic-sounders" on data.gov) —as well as places ("Hawaii," "Canada," "Europe"), and many different kinds of entities and concerns ("mammal," "school," "ocean," "ships," "tree rings," "wildlife," "reef," "borehole," "webcam," "oil spill," "day patient," "earnings," "lithosphere," "bus," "prison," "milk," "woodland," "carbon," "floods," "death"). Data portal tags indicate how the byproducts of quantification and data collection involved in coordinating and managing collective life are labelled and presented for broader public consumption. In complement to studies looking at broader semantic patterns in metadata (Adel Rezk et al., 2017), there is also perhaps a case to be made for reading these specific kinds of granular traces alongside institutional histories and arrangements for making things legible and intelligible in particular ways (Gray, 2018b).

Data portal metadata can also provide insights into how data is being put to work by different actors. For example, metadata from data.gov.uk's app catalog can be repurposed in order to explore relations between tags and apps, giving us a picture of how public sector data is used to generate different ways of knowing, experiencing and interacting with data (Figure 13). Many of these focus on location-based personalization (e.g., find a doctor, find childcare, find a school) and decision-making and choice (compare care homes, check food and hygiene). Towards the center of the graph are apps that combine different datasets and indicators in order to operationalize notions such as "walkability," "liveability" and

Figure 13. *Bipartite network showing apps and user-generated tags associated with them from* *Data.gov.uk* *"app catalog." Created using the Gephi software, spatialized using the ForceAtlas2 graph layout algorithm (Jacomy et al., 2014) and annotated through qualitative analysis.*

"pleasantness"—suggesting a kind of combinatorial analytic creativity, as well as how data is used for personal environmental optimization.

Sometimes data requests and their responses are made public, which enables one to see when they are successful (e.g., when institutions reply with data being requested) and sometimes not (e.g., when data is not generated or when it cannot be published). Data portal mechanisms can facilitate encounters between citizens and public servants, surfacing different understandings and assumptions about what data portals are for, what can be done on them, and what kinds of data one might expect institutions to gather and open up. Whether they are successful or not, requests and suggestions can indicate what data portal visitors come looking for and why, what they hope or expect to find, what is and is not counted as public data, what kinds of information public bodies do and do not collect and make available, and what kinds of questions it is and is not possible to answer with public sector data. Unsuccessful requests include those for data which can be more specific than that generated by institutions (e.g., "accidents with injuries as a result of an activity with a horse in the Netherlands"), with different methods, starting points, coverage or time spans. Data requests probe the limits of what can be made public (e.g., "addresses where a firearm is available"), as well as the limits of what kinds of information institutions routinely and systematically gather as part of their defined tasks (e.g., "list of municipal councillors who have split from their local political party").

Figure 14. *Data not found, a dataset of datasets that were sought but not found on data portals around the world. Available at: http://datanotfound.jwyg.org/.*

Compiling unsuccessful data requests from across portals surfaces a plethora of interests, concerns, curiosities, and queries for data that does not exist or which is not made public (Figure 14).[34] Rather than assuming a kind of "data universalism" (Milan and Treré, 2019), these unsuccessful attempts to obtain data through portals highlight some of the different ways in which data comes to matter to different people in different situations. While the United Nations talks of "data gaps,"[35] how do we establish what is to be considered absent? Mimi Ọnụọha's "Library of Missing Datasets" draws attention to "things that have been excluded in a society where so much is collected," and invites us to reflect on what is counted, what is not, and who decides.[36] In the spirit of feminist STS researcher Susan Leigh Star's unfolding of relations between "lived experience," "technologies" and "silences" (Star, 2007), what is considered missing data, absent data, a "data gap" or a "data void" can be taken as an open-ended question to be addressed in relational and situational terms.[37] Data portals may thus serve as an entry-point not only for finding official datasets, but also for the empirical elaboration of missingness and the questions and lines of inquiry which cannot be addressed by current arrangements.

4. Conclusion

This article explores what research on data portals can draw from device-sensitive research approaches—including to examine what data portals produce, their makers, users, parts, processes, and politics as they

---

[34] http://datanotfound.jwyg.org/.
[35] For example, https://news.un.org/en/story/2014/08/476242 and https://press.un.org/en/2015/sgsm17208.doc.htm.
[36] https://mimionuoha.com/the-library-of-missing-datasets.
[37] This is further elaborated in chapter five of my *Data Worlds* book, under contract with MIT Press.

Table 1. *Illustrative analytical scenarios for studying data portals as online devices*

Entrypoint	Method	Illustrative research question
Data portal interfaces	Reverse wireframing	How are data portal users configured?
	Web histories	How have data portals changed over time?
	Feature analysis	How do data portals organize participation?
	Data topic analysis	How are data topics prioritized and displayed?
Data portal software	Commit analysis	How has data portal software changed over time?
	Account analysis	Who is making and using data portal software?
	Fork analysis	How is data portal software modified?
	Extension analysis	How is data portal software customized and extended?
Data portal metadata	Schema analysis	How are public data relations envisaged and modeled?
	API analysis	How is official data being made programmable?
	Category analysis	Which data topics are more and less prevalent?
	Dataset analysis	What kinds of concerns are legible in public data?
	Tag analysis	How is data folksonomized by portal contributors?
	App analysis	How does data use script citizen-state-market relations?
	Requests analysis	What kinds of data is sought (and not found)?

render informational resources of public institutions as more widely re-usable digital data. The sections above have outlined and illustrated three approaches to studying data portals as online devices (Table 1): firstly, by attending to data portal interfaces, including reverse wireframing and web histories; second, by examining data portal software, including code repositories and software extensions; and third, by looking at data portal metadata, including domain models as well as metadata on topics, tags, organizations, apps, and data requests.

As well as supporting the study of individual data portals these approaches may surface conventions, features, and libraries which are shared across multiple portals, whether local, regional, national, or transnational in scope. Taking cue from device-sensitive research, these three approaches can help to surface the various actors, materials, and processes which are involved in operationalizing open government data commitments, as well as associated tensions, frictions, instabilities involved in the "rawification" of public sector information (Denis and Goëta, 2017).

Data portals are not just technical systems to operationalize pre-existing policy commitments, but socio-technical assemblages which gather heterogenous actors to take part in data-mediated transparency, participation, and innovation. They do many different kinds of things, for many different kinds of actors with many different kinds of objectives. Their material organization can be unpacked and understood as the result of an accumulation of decisions, negotiations, circumstances, and contingencies. The study of portal interfaces, software, and metadata may contribute to relational approaches to the configuration of state institutions in society (Thelen et al., 2017); research that aims to situate data practices in society (Leurs, 2017; Rettberg, 2020); as well as "post-instrumental" understandings of how online devices are increasingly equipped with capacities to take part (Marres, 2012), rather than just serving as part of the scenery. Attending to data portals as sites of friction (Tsing, 2011), for example, in failed data portal requests; and translation (Callon, 1984), for example, as a collaborative knowledge tool is adapted for institutional publishing needs, may contribute to de-instrumentalizing data portals and being alive to their politics, dynamics, social lives and possibilities as online devices. As well as serving to commensurate (Espeland and Stevens, 1998) wide-ranging informational resources and publish them as data in a way that is amenable to various forms of quantification (Espeland and Stevens, 2008), data portals draw on a range of different practices and action repertoires to make data available, from public-facing features to support searching, browsing, discussion, sharing, and requesting, to institutional mechanisms for gathering and harmonizing metadata, indexing datasets and monitoring activity.

How do these approaches modify understanding of the prospects of data portals? May they be viewed as agents of data colonialism (Couldry and Mejias, 2019) or surveillance capitalism (Zuboff, 2019), extending the platforms and logics of big tech ever more into public institutions, workplaces, and everyday life through a kind of information subsidy to corporates? As part of longer histories of colonialism (Hoyng, 2021) and economic discipline (Roberts, 2011), involved in both the making of particular kinds of states and state-market-citizen relations, as well as emerging arrangements of planetary scale computation (Bratton, 2016)? For critical data studies (Dalton and Thatcher, 2014; Dalton et al., 2016; Iliadis and Russo, 2016), attending to the details of portal interfaces, software, and metadata may be used to empirically elaborate how commodification unfolds. This includes assessing how may fail to deliver on their aspirations—for example, by highlighting "missing users" (Heald, 2012), limited and non-diverse participation, short-lived projects, regressive redistribution, or how apps and outcomes often serve marketization and commercialization of public data rather than centering issues and under-served communities. Studying data portals as online devices may show how some aims and interests are much better served than others, as well as helping to situate data portals in relation to troubling tendencies and agendas.

Will data portals continue to be created and maintained? And if so, how might they be done differently? For example, could one rethink or remake data portals along the lines of recent work on data feminism (D'Ignazio and Klein, 2020), critical data practices (Bounegru and Gray, 2021) or undoing data-driven optimization (Powell, 2021)? Could they serve as sites for societal issue articulation as well as official accounts, affected communities rather than data businesses, data commons rather than data capitalism, public–common partnerships rather than public-private part-nerships?[38] Rather than thinking of data portals instrumentally in relation to more liberal, minimal conventions of opening up government information (Gray, 2018a), could they be envisaged as sites of social learning (Williams et al., 2005), experiments in participation (Lezaun et al., 2016), or as test sites (Marres and Stark, 2020) around the changing roles of institutional data in society? What could a feminist, ecological, decolonial, new economy or post-platform data portal look like? As well as sharpening critique, device-oriented approaches may also surface how data portals do not go according to plan or may give rise to unexpected encounters and interactions, such as through unsuccessful and unanswerable data requests. Perhaps in the details of such materials may be found clues as to what portals might look like if they served not only to serve up official representations and responses to requests, but also were alive to questions, curiosities, and concerns which could inform other kinds of data infrastructures and arrangements (Gray et al., 2018; Gray, 2018b). Close attention to the details of interfaces, software, and metadata, may provide inspiration for how these might be repurposed, re-appropriated, and recomposed with other values and sensibilities, should circum-stances arise.

---

[38] https://www.common-wealth.co.uk/reports/public-common-partnerships-building-new-circuits-of-collective-ownership.

with other colleagues at the Public Data Lab. Thanks to Sebastian Stros, Elena Wüllhorst, Khizer Sajid for their contributions to exploratory research which informed this paper, undertaken as part of the King's Undergraduate Research Fellowship (KURF) scheme in Summer 2021. Finally, I'm most grateful for feedback and encouragement from participants at workshops and conferences where earlier versions of this work were presented, including a keynote at the Digital Methods Winter School 2017 at the University of Amsterdam; the *Data Publics* conference at Lancaster University; the *Open Data Movements in the Age of Big Data Capitalism* workshop at the University of Westminster; the *International Conference on Public Policy 2017* at the National University of Singapore; the *Outnumbered! Statistics, Data and the Public Interest* workshop at the University of Cambridge; a keynote at *Deuxième Journée Open Data: Culture des données et données pour la culture*, at the Université Paris Nanterre, Paris; the *Media and Publics Conference* at Roskilde University; a keynote at the *Citizen Friendly Data Communication* conference, Tilburg University; and a talk as part of the *Seminar of the Society for the Social Sciences of Quantification (SSSQ)* in Paris.

References

Adel Rezk M, **Ojo A and Hassan IA** (2017) Mining governmental collaboration through semantic profiling of open data catalogues and publishers. In LM Camarinha-Matos, Afsarmanesh H and Fornasiero R (eds), *Collaboration in a Data-Rich World*, Cham, Switzerland: Springer, pp. 253–264.

Alexopoulos C, **Diamantopoulou V and Charalabidis Y** (2017) Tracking the evolution of OGD portals: A maturity model. In Janssen M, Axelsson K, Glassey O, Klievink B, Krimmer R, Lindgren I, Parycek P, Scholl HJ, Trutnev D Trutnev D (eds), *Electronic Government*. Cham: Springer International Publishing, pp. 287–300.

Andersen CU and Pold S (eds) (2011) *Interface Criticism: Aesthetics beyond Buttons*. Aarhus, Denmark: Aarhus University Press.

Andersen CU and Pold S (2018) *The Metainterface: The Art of Platforms, Cities, and Clouds*. Cambridge, MA; London, England: The MIT Press.

Barry A (2013) Transparency as a political device. In Akrich M, Barthe Y, Muniesa F and Mustar P (eds), *Débordements: Mélanges Offerts à Michel Callon*. Paris: Presses des Mines, pp. 21–39.

Bates J (2012) "This is what modern deregulation looks like": Co-optation and contestation in the shaping of the UK's open government data initiative. *The Journal of Community Informatics 8*. Available at https://doi.org/10.15353/joci.v8i2.3038 (accessed 7th July 2022).

Birchall C (2015) 'Data.Gov-in-a-box': Delimiting transparency. *European Journal of Social Theory 18*, 185–202.

Boucquey N, **Martin KS**, **Fairbanks L**, **Campbell LM and Wise S** (2019) Ocean data portals: Performing a new infrastructure for ocean governance. *Environment and Planning D: Society and Space 37*, 484–503.

Bounegru L (2019) *News Devices: How Digital Objects Participate in News and Research (University of Groningen)*. University of Groningen.

Bounegru L and Gray J (eds) (2021) *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press.

Bratton BH (2016) *The Stack: On Software and Sovereignty*. Cambridge, MA: MIT Press.

Bruno I, **Jany-Catrice F and Touchelay B** (eds) (2016) *The Social Sciences of Quantification: From Politics of Large Numbers to Target-Driven Policies*. New York: Springer.

Burns R and Wark G (2020) Where's the database in digital ethnography? Exploring database ethnography for open data research. *Qualitative Research 20*, 598–616.

Callon M (1984) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review 32*, 196–233.

Callon M, **Millo Y and Muniesa F** (2007) *Market Devices*. Chichester: Wiley.

Chun WHK (2013) *Programmed Visions: Software and Memory*. Cambridge, MA: MIT Press.

Couldry N and Mejias UA (2019) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford, CA: Stanford University Press.

Cramer F and Fuller M (2008) Interface. In Fuller M (ed.), *Software Studies: A Lexicon*. Cambridge, MA: MIT Press.

Currie M (2020) Data as performance – Showcasing cities through open data maps. *Big Data & Society 7*, 2053951720907953.

D'Ignazio C and Klein L (2020) *Data Feminism*. Cambridge, MA: MIT Press.

Dalton CM, **Taylor L and Thatcher J** (alphabetical) (2016) Critical data studies: A dialog on data and space. *Big Data & Society*, 3, 2053951716648346.

Dalton CM and Thatcher J (2014) What does a critical data studies look like, and why do we care? *Society & Space*. Available at https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care (accessed 7th July 2022).

Dantec CAL and DiSalvo C (2013) Infrastructuring and the formation of publics in participatory design. *Social Studies of Science* 43, 241–264.

Davies T (2021) The pressure on portals: An hourglass approach. *Data Portals and Citizen Engagement*. Available at https://dataportals.pubpub.org/pub/qk1untv2/release/2 (accessed 7th July 2022).

de Mourat R, **Ricci D and Latour B** (2020) How does a format make a public? In Eve MP and Gray J (eds), *Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access*. Cambridge, MA: MIT Press.

Denis J and Goëta S (2017) Rawification and the careful generation of open government data. *Social Studies of Science 47*, 604–629.

Desrosières A (2002) In (Naish C, trans.) (eds), *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

Dourish P (2007) Seeing like an interface. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces*. New York, NY: ACM, pp. 1–8.

Espeland WN and Stevens ML (1998) Commensuration as a social process. *Annual Review of Sociology 24*, 313–343.

Espeland WN and Stevens ML (2008) A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie 49*, 401–436.

Flyverbom M (2016) Transparency: Mediation and the management of visibilities. *International Journal of Communication 19*, 110–122.

Frankel C, **Ossandón J and Pallesen T** (2016) Markets for collective concerns. In *Presented at the EASST 2014 Conference (European Association for the Study of Science and Technology)*. https://research.cbs.dk/en/publications/markets-for-collective-concerns

Galloway AR (2012) *The Interface Effect*. Cambridge, UK; Malden, MA: Polity Press.

Gray J (2014) Towards a genealogy of open data. In *Presented at the European Consortium for Political Research (ECPR) General Conference 2014, University of Glasgow*. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2605828 (accessed 7th July 2022).

Gray J (2017) Digital methods and public policy: Tracing networks, assemblages and devices. In *Presented at the International Conference on Public Policy, National University of Singapore* https://doi.org/10.5281/zenodo.7741522.

Gray J (2018a) Quand les mondes de données sont redistribués: Open data, infrastructures de données et démocratie. *Statistique et Société 5*, 29–34.

Gray J (2018b) Three aspects of data worlds. *Krisis: Journal for Contemporary Philosophy*. Available at http://krisis.eu/three-aspects-of-data-worlds/ (accessed 7th July 2022).

Gray J, **Gerlitz C and Bounegru L** (2018) Data infrastructure literacy. *Big Data & Society 5*, 1–13.

Gurstein MB (2011) Open data: Empowering the empowered or effective data use for everyone? *First Monday 16*. Available at http://firstmonday.org/ojs/index.php/fm/article/view/3316 (accessed 7th July 2022).

Hansen HK and Flyverbom M (2015) The politics of transparency and the calibration of knowledge in the digital age. *Organization 22*, 872–889.

Harvey P, **Reeves M and Ruppert E** (2013) Anticipating failure: Transparency devices and their effects. *Journal of Cultural Economy 6*, 294–312.

Heald D (2012) Why is transparency about public expenditure so elusive? *International Review of Administrative Sciences 78*, 30–49.

Helmond A (2015) The platformization of the web: Making web data platform ready. *Social Media + Society 1*, 2056305115603080.

Hirschman D and Berman EP (2014) Do economists make policies? On the political effects of economics. *Socio-Economic Review 12*, 779–811.

Hookway B (2014) *Interface*. Cambridge, Massachusetts: MIT Press.

Hoyng R (2021) From open data to "grounded openness": Recursive politics and postcolonial struggle in Hong Kong. *Television & New Media 22*, 703–720.

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society 3*, 2053951716674238.

Jacomy M, **Venturini T**, **Heymann S and Bastian M** (2014) ForceAtlas2, a continuous graph layout algorithm for Handy network visualization designed for the Gephi software. *PLoS One 9*, e98679.

Jasanoff S and Kim S-H (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press.

Kelty CM (2005) Geeks, social imaginaries, and recursive publics. *Cultural Anthropology 20*, 185–214.

Kelty CM (2008) *Two Bits: The Cultural Significance of Free Software: The Cultural Significance of Free Software and the Internet*. Durham, NC: Duke University Press.

Law J and Ruppert E (2013) The social life of methods: Devices. *Journal of Cultural Economy 6*, 229–240.

Lehtiniemi T and Ruckenstein M (2019) The social imaginaries of data activism. *Big Data & Society 6*, 2053951718821146.

Leonard S (2012) The Fog of More. *The New Inquiry.* Available at https://web.archive.org/web/20160829085143/; http://thenewinquiry.com/essays/the-fog-of-more/ (accessed 7th July 2022).

Leonelli S (2021) Data science in times of pan(dem)ic. *Harvard Data Science Review 3.* https://doi.org/10.1162/99608f92.fbb1bdd6.

Leurs K (2017) Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review 115*, 130–154.

Lezaun J, **Marres N and Tironi M** (2016) Experiments in participation. In Felt U, Fouche R, Miller CA and Smitt-Doer L (eds), *Handbook of Science and Technology Studies*, 4th Edn. Cambridge: MIT Press.

Light B, **Burgess J and Duguay S** (2018) The walkthrough method: An approach to the study of apps. *New Media & Society 20*, 881–900.

Liu A (2004) *The Laws of Cool: Knowledge Work and the Culture of Information*, 2nd Edn. Chicago: University of Chicago Press.

Longo J (2011) OpenData: Digital-era governance thoroughbred or new public management Trojan horse? *Public Policy and Governance Review 2*, 38–51 Available at http://ppgreview.ca/past-issues/volume-2-issue-2-2011/ (accessed 7th July 2022).

Machado V, **Mantini G**, **Viterbo J**, **Bernardini F and Barcellos R** (2018) An instrument for evaluating open data portals. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1–10 https://doi.org/10.1145/3209281.3209370.

Manovich L (2008) Data visualization as new abstraction and as anti-sublime. In Hawk B, Rieder DM and Oviedo O (eds), *Small Tech: The Culture of Digital Tools*. Minneapolis: University of Minnesota Press.

Marres N (2011) The costs of public involvement: Everyday devices of carbon accounting and the materialization of participation. *Economy and Society 40*, 510–533.

Marres N (2012) *Material Participation: Technology, the Environment and Everyday Publics*. London: Palgrave Macmillan.

Marres N (2017) *Digital Sociology: The Reinvention of Social Research*. London: Polity Press.

Marres N and Lezaun J (2011) Materials and devices of the public: An introduction. *Economy and Society 40*, 489–509.

Marres N and Stark D (2020) Put to the test: For a new sociology of testing. *The British Journal of Sociology 71*, 423–443.

Milan S and Treré E (2019) Big data from the south(s): Beyond data universalism. *Television & New Media 20*, 319–335.

Morozov E (2013) The meme hustler: Tim O'Reilly's crazy talk. *The Baffler 22*. Available at https://thebaffler.com/salvos/the-meme-hustler (accessed 7th July 2022).

Neumaier S, **Umbrich J and Polleres A** (2016) Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality 8*(1), 2.

Neurath M and Kinross R (2009) *The Transformer: Principles of Making Isotype Charts*. London: Hyphen Press.

Nikiforova A and McBride K (2021) Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics 58*, 101539.

Nogueras-Iso J, **Lacasta J**, **Ureña-Cámara MA and Ariza-López FJ** (2021) Quality of metadata in open data portals. *IEEE Access 9*, 60364–60382.

O'Reilly T (2010) Government as a platform. *Innovations: Technology, Governance, Globalization 6*, 13–40.

Pold S (2005) Interface realisms: The Interface as aesthetic form. *Postmodern Culture 15*. https://doi.org/10.1353/pmc.2005.0013

Powell AB (2021) *Undoing Optimization: Civic Action in Smart Cities*. New Haven: Yale University Press.

Ratner H and Ruppert E (2019) Producing and projecting data: Aesthetic practices of government data portals. *Big Data & Society 6*, 2053951719853316.

Rettberg JW (2020) Situated data analysis: A new method for analysing encoded power relationships in social media platforms and apps. *Humanities and Social Sciences Communications 7*, 1–13.

Rieder G (2018) Tracing big data imaginaries through public policy: The case of the European Commission. In Sætnan AR, Schneider I and Green N (eds), *The Politics and Policies of Big Data: Big Data, Big Brother?* London: Routledge.

Roberts A (2011) *The Logic of Discipline: Global Capitalism and the Architecture of Government*. Oxford: Oxford University Press.

Rogers R (2004) *Information Politics on the Web*. Cambridge, MA: MIT Press.

Rogers R (2013) *Digital Methods*. Cambridge, MA: MIT Press.

Rogers R (2017) Doing web history with the internet archive: Screencast documentaries. *Internet Histories 1*, 160–172.

Ruppert E (2015) Doing the transparent state: Open government data as performance indicators. In Rottenburg R, Merry SE, Park S-J and Mugler J (eds), *A World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge, UK: Cambridge University Press, pp. 127–150.

Ruppert E (2018) *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data*. Rotterdam: Erasmus Universiteit Rotterdam.

Ruppert E, **Law J and Savage M** (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society 30*, 22–46.

Shadbolt N, **O'Hara K**, **Berners-Lee T**, **Gibbins N**, **Glaser H**, **Hall W and Schraefel MC** (2012) Linked open government data: Lessons from Data.gov.uk. *IEEE Intelligent Systems 27*, 16–24.

Slee T (2012) Seeing Like a Geek. Available at http://crookedtimber.org/2012/06/25/seeing-like-a-geek/ (accessed 3 December 2018).

Star SL (2007) Five questions. In Olsen JKB and Selinger E (eds), *Philosophy of Technology*. New York: Automatic Press.

Suchman LA (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd Edn. Cambridge: Cambridge University Press.

Taylor C (2002) Modern social imaginaries. *Public Culture 14*, 91–124.

Thelen T, **Vetters L and von Benda-Beckmann K** (eds) (2017) *Stategraphy: Toward a Relational Anthropology of the State*. Oxford, New York: Berghahn Books.

Tkacz N (2012) From open source to open government: A critique of open politics. *Ephemera: Theory and Politics in Organization 12*, 386–405.

Tsing AL (2011) *Friction: An Ethnography of Global Connection*. Princeton, NJ: Princeton University Press.

Tupasela A, **Snell K and Tarkkala H** (2020) The Nordic data imaginary. *Big Data & Society 7*, 2053951720907107.

Weltevrede E, **Helmond A and Gerlitz C** (2014) The politics of real-time: A device perspective on social media platforms and search engines. *Theory, Culture & Society 31*, 125–150.

Williams R, **Stewart J and Slack R** (2005) *Social Learning in Technological Innovation – Experimenting with Information and Communication Technologies*. Cheltenham, UK: Edward Elgar Publishing.

Woolgar S (1990) Configuring the user: The case of usability trials. *The Sociological Review 38*, 58–99.

Zuboff PS (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frpontier of Power*. London: Profile Books.